



TRABAJO DE FIN DE GRADO

<<Análisis estadístico-descriptivo de rasgos pragmáticos para la atribución de autoría en español>>

Autora: Ana Sánchez Sánchez

Tutores: Nuria Campos Carrasco y Mario Crespo Miguel

Doble Grado en Filología Hispánica y en Lingüística y Lenguas Aplicadas

Curso Académico 2018-2019

Fecha de presentación 31/05/2019



FACULTAD DE FILOSOFÍA Y LETRAS

Índice:

1. Resumen	3
2. Introducción	4
3. La Lingüística Forense como disciplina de estudio	6
3.1. Definición	6
3.2. Ámbitos	8
3.3. Origen	10
3.4. Situación en España y perspectivas actuales	11
4. La atribución de autoría. Conceptos generales	14
5. El registro coloquial como instrumento de análisis forense en español	17
5.1. El registro coloquial en español	17
5.1.1 El registro coloquial y la conversación	24
5.2. Marcas identificativas del registro coloquial	28
5.2.1. Marcas morfosintácticas	28
5.2.2. Marcadores pragmático-discursivos	31
6. Desarrollo práctico	41
6.1. Hipótesis de trabajo y metodología	41
6.1.1. Tests de hipótesis y prueba Chí cuadrado	43
6.2. Análisis del corpus	44
6.2.1. Análisis Chí cuadrado	45
6.2.2. Tabla de error	54
6.2.3. Análisis perfiles lingüísticos	55
7. Conclusión	63
8. Bibliografía	64

1. Resumen:

En esta investigación se ha llevado a cabo una combinación de trabajo teórico y práctico. Así, hemos tratado de realizar una atribución de autoría de textos dubitados basándonos en la teoría y la metodología más comúnmente empleadas en el ámbito de la identificación de autores. Junto a esto, el objetivo pragmalingüístico de nuestro trabajo ha sido indagar sobre la posibilidad de singularizar patrones lingüísticos de diferentes perfiles de hablantes de registro coloquial a partir de una serie de marcas sintácticas y pragmáticas contribuyendo así con mayor precisión a la atribución autoría de textos dubitados en español. Todos los textos empleados en nuestro estudio han sido tomados del corpus ForenUca, desarrollado por el Instituto de Lingüística Aplicada (ILA) de la Universidad de Cádiz en los que se recopila una serie de textos producidos en diferentes medios de comunicación (mensajería corta, email, redes sociales). Concretamente, los textos que hemos seleccionado del corpus han sido los producidos en la aplicación de mensajería instantánea wasap. De los 40 textos que hemos analizado, 5 de ellos son de autoría indubitada (por tanto, conocemos su autor) y otros 5 son de autoría que hemos denominado dubitada, aunque sabemos que pertenecen al mismo autor (esta decisión fue tomada para comprobar que nuestra herramienta estadística de observación (UAM Corpus Tool) era fiable para realizar el trabajo). A partir de este corpus, hemos analizado y observado si el uso de las marcas coloquiales se mantenía entre los mismos autores o no, y si estas se comportaban de la misma manera entre los textos dubitados y el general. Los textos restantes fueron agrupados bajo el nombre “general” y nos sirvieron para comprobar el comportamiento de los autores con respecto a los demás hablantes. A partir de estos textos y mediante el citado programa informático, hemos comprobado si las marcas sintácticas y pragmáticas que hemos seleccionado previamente como indicadoras de registro coloquial eran relevantes o no entre hablantes que utilizan un registro coloquial y con el entorno wasap como escenario de comunicación. Una vez obtenidos los resultados, hemos concluido la autoría de los textos y hemos realizado el perfil lingüístico de los 5 hablantes. Por último, hemos concluido que en el entorno comunicativo de wasap es posible identificar marcas asociadas con el registro coloquial en español y, por tanto, los resultados obtenidos en el programa utilizado como herramienta han sido exitosos para la investigación.

Palabras clave: lingüística forense, pragmática, registro coloquial, atribución de autoría, perfil lingüístico, test chí cuadrado.

In this research a combination of theoretical and practical work has been carried out. In this way, we have tried to make an attribution of authorship of doubtful texts based on the theory and methodology most commonly used in the field of identification of authors. Together with this, the pragmatic and linguistic objective of our work has been to investigate the possibility of singularizing linguistic patterns of different profiles of speakers of colloquial register from a series of syntactic and pragmatic marks, to contribute with greater precision to the attribution of authorship of doubtful texts in Spanish. All the texts used in our study have been taken from the corpus ForenUca, developed by the Institute of Applied Linguistics (ILA) of the University of Cadiz which compiles a series of texts produced in different media (short messaging, email, social networks). Specifically, the texts we have selected from the corpus have been produced in the application of instant messaging whatsapp. Of the 40 texts we have analysed, 5 of them are of undoubted authorship (therefore, we know its author) and other 5 are of authorship that we have denominated doubtful, although we know that they belong to the same author (this decision was taken to check that our statistical observation tool (UAM Corpus Tool) was reliable to carry out the study). From this corpus, we have analysed and observed whether the use of colloquial marks was stable in the texts of the same author or not, and if these behaved in the same way between the doubtful texts and the general group. The remaining texts were grouped under the name "general" in order to check the behaviour of the authors compared to the other speakers. From these texts and with the results provided by the mentioned computer program, we have checked if the syntactic and pragmatic marks that we have previously selected as indicators of colloquial register were relevant or not between speakers that use a colloquial register and with the whatsapp environment as a communication scenario. Once the results were obtained, we concluded the authorship of the texts and made the linguistic profile of the 5 speakers. Finally, we have concluded that in the communicative environment of whatsapp it is possible to identify marks associated with the colloquial register in Spanish and therefore, the results obtained in the program have been successful for the research.

Keywords: forensic linguistics, pragmatics, colloquial register, authorship attribution, linguistic profile, chi square test.

2. Introducción:

Nuestra investigación se nutre de los conocimientos de una aún incipiente lingüística forense para combinarlos con los de la pragmática generando así un modelo propio de análisis

adaptado al español. En este trabajo conjunto, nuestro objetivo general es doble: por una parte, investigar sobre la posibilidad de establecer patrones lingüísticos que singularicen perfiles de uso en registro coloquial; por otra, en el ámbito de la lingüística forense este tipo de patrones deberá contribuir a la identificación de tipos de hablantes en textos dubitados anónimos. Tomando como punto de partida marcas sintácticas y pragmáticas que la literatura científica relaciona con el empleo del registro coloquial por parte del hablante de español, identificaremos la frecuencia de uso de las mismas en un corpus de producciones. Además, trataremos de establecer qué variables (edad, sexo y formación) inciden en el uso de estas marcas. También exploraremos la posibilidad, si cabe, de dibujar perfiles siempre que unas determinadas variables activen las mismas marcas de manera recurrente. El desarrollo futuro de nuestro estudio consistirá en ampliar el corpus de producciones para poder refutar o refrendar las conclusiones preliminares. Es decir, al ampliar el corpus, podremos observar si los resultados son similares a los obtenidos en nuestra investigación y por tanto se confirman las conclusiones preliminares de este estudio o, si por el contrario, los datos muestran otro comportamiento diferente que nos haga sacar otras conclusiones. También sería interesante identificar las posibles diferencias de uso de marcas en distintos entornos comunicativos con el mismo registro coloquial (por ejemplo, wasap frente a producción oral espontánea presencial).

El primer capítulo consistirá en presentar la lingüística forense como disciplina de estudio. Para ello, haremos un breve recorrido sobre las distintas definiciones que se le ha dado, hablaremos sobre los ámbitos de estudio y el origen de la misma y concluiremos con la situación actual de la lingüística forense en España.

Una vez conocido esto, pasaremos a hablar sobre la atribución de autoría. Conocer las bases teóricas de esta técnica de investigación será fundamental para desarrollar nuestro trabajo pues, como ya hemos comentado anteriormente, nuestro objetivo es establecer patrones lingüísticos para la identificación de hablantes.

El siguiente capítulo desarrollará nociones básicas sobre el uso del registro coloquial como instrumento de análisis forense español. Intentaremos justificar que, basándonos en el registro coloquial y en sus producciones lingüísticas, podremos identificar una serie de marcas como potencialmente distintivas de hablantes de español. Una vez seleccionadas las marcas, comenzaremos el desarrollo práctico de la investigación. Presentaremos y analizaremos los resultados obtenidos del programa UAM Corpus Tool y sacaremos unas conclusiones. Finalmente, en el último capítulo concluiremos nuestra investigación.

3. La lingüística forense como disciplina de estudio:

3.1. Definición.

3.2. Ámbitos.

3.3. Origen.

3.4. Situación en España y perspectivas actuales.

En este capítulo pasaremos a hablar de la lingüística forense. Discutiremos sobre las diferentes definiciones que se le han otorgado a la disciplina y adoptaremos una de las dos perspectivas en las que se enmarca para poder comenzar con nuestro estudio. Además, comentaremos las distintas áreas de la lingüística forense y explicaremos la labor que ejerce el lingüista forense en el ámbito legal y judicial. Por otro lado, nos remontaremos a su origen hasta llegar a exponer la situación actual de la lingüística forense en España.

3.1 Definición:

La lingüística forense es una disciplina de estudio multidisciplinar e interdisciplinar en la que participan numerosos profesionales y expertos (tanto lingüistas como otros), que, aunque sea reciente en España, posee una amplia trayectoria en el mundo anglosajón (Bernal, Riveiros, Garayzábal, 2014). Se le considera una aplicación de la lingüística y, a pesar de que no haya una definición consensuada de la disciplina, esta se puede enmarcar en dos perspectivas diferentes. Por un lado, la lingüística forense se puede definir mediante una perspectiva restrictiva que consiste en la utilización de la lengua y las evidencias lingüísticas como parte de las pruebas para investigar un crimen o, por otro lado, desde una perspectiva más amplia en la que el término se define como la interfaz entre Lengua y Derecho (definición que toma la IAFL de la cual hablaremos más adelante): así lo dijo Gibbons en 1999. Este autor presenta la falta de consenso entre los expertos lingüistas y juristas para definir el término y comenta las dos perspectivas posibles para estudiar la lingüística forense: “There is no consensus among linguistics and legal experts as to a definition of the term forensic linguistics. Some adhere to a narrow definition such as the use of linguistic techniques to investigate crimes in which language data constitute part of the evidence. Others subscribe to a broad interpretation of forensic linguistics as the study of the intersection between language and the law” (Bernal, Riveiros, Garayzábal, 2014, p. 22).

Según el Diccionario Legal de Black de 1990, la lingüística forense es “una técnica relacionada con una profunda y detallada evaluación de las características lingüísticas de un

texto, incluyendo gramática, sintaxis, ortografía, vocabulario y fraseología, que se lleva a cabo mediante la comparación de textos dubitados (de autoría desconocida) e indubitados (de autoría conocida) en un intento de descubrir idiosincrasias peculiares de autoría a fin de determinar si los autores pueden o no ser el mismo” (Bernal, Riveiros, Garayzábal, 2014, p. 22). Para David Crystal (1987), sin embargo, se trata de una aplicación de la estilística cuando alguien es sospechoso de haber incumplido la ley. Posteriormente, en 1997, rectificó esa afirmación y definió la lingüística forense como “el uso de las técnicas lingüísticas para investigar delitos en los que los datos relacionados con el lenguaje forman parte de las pruebas, tales como el uso de criterios léxicos o gramaticales para autenticar declaraciones policiales” (Bernal, Riveiros, Garayzábal, 2014, p. 21). Como podemos ver, esta última definición puede considerarse dentro de la perspectiva restrictiva. Otros autores como Olsson (2010) la define como “the study, analysis and measurement of language in the context of crime, judicial procedure, or disputes in law, including the preparation and giving of written and oral evidence” (Bernal, Riveiros, Garayzábal, 2014, p. 22). Asimismo, el centro de lingüística forense de la Universidad de Aston, en Birmingham defiende que “Forensic Linguistics can be fairly characterised as taking linguistic knowledge, methods and insight, and applying these to the forensic context of law, investigation, trial, punishment and rehabilitation” (Bernal, Riveiros, Garayzábal, 2014, p. 23). Mireia Hernández Esteban (2016) recoge en su libro *Lingüística Forense Básica, Metodología para la atribución de autoría de un texto* las definiciones de varios profesores. Primeramente, José Luis Cifuentes Honrubia, profesor de la Universidad de Alicante considera que la lingüística forense es un punto de encuentro entre la Lingüística y el Derecho, por lo que también se atiene a la perspectiva abierta. Además, añade que es una disciplina lingüística que incluye áreas de investigación estrechamente unidas al lenguaje administrativo, jurídico y judicial. Por otro lado, los profesores Viñals y Puente (2009) definen a la lingüística forense como una “investigación judicial del discurso donde se incluye la sociolingüística (pragmalingüística, análisis del discurso, retórica, lingüística textual, computarizada etc.) relacionándose con la antropología, historia, sociología, psicología, pedagogía, comunicación informativa etc., lo que permite obtener un perfil de personalidad del que se expresa, interpretar o decodificar adecuadamente un mensaje en su contexto, revelando la verdadera intencionalidad del autor, reconocer las estrategias lingüísticas empleadas en el texto, así como de su influencia (manipulación) en el receptor” (p. 12). La autora también aporta su propia definición y afirma que la lingüística forense es una herramienta de trabajo que aporta indicios que ayudan a resolver problemas de atribución de autoría. La definición que da Gerald McMenamim (2002) aporta que la

lingüística forense es “the application of linguistic knowledge to a particular social setting, namely the legal *forum* (from which the word *forensic* is derived)” (p. 12). Este autor también ubica a la disciplina dentro de la perspectiva amplia¹ y además la relaciona con otros campos de estudio, entre ellos la teoría de los actos de habla². No nos es fácil atenernos a una sola definición de la disciplina, pues todas las mencionadas anteriormente añaden matices de valor que ayudan a entender con mayor claridad el objeto de estudio de la lingüística forense. Sin embargo, tras nuestro breve recorrido por las diferentes definiciones, está claro que la lingüística forense tiene como objetivo fundamental la resolución de problemas en la sociedad mediante el análisis del lenguaje y, de este modo, podemos enmarcarla entre las ramas de la lingüística aplicada (Ramírez Salado, 2017). En nuestro caso (al igual que Ramírez Salado, 2017), vamos a optar por entender la lingüística forense como la disciplina que, mediante el empleo de estrategias lingüísticas, investiga delitos en los que el lenguaje forma parte de las pruebas. Hemos decidido decantarnos por esta perspectiva porque es la que está ligada al origen del término *forensic linguistics* acuñado por Svartvik en su análisis del caso Evans (del cual hablaremos más adelante).

3.2. Ámbitos:

Como hemos comentado anteriormente, la lingüística forense es una disciplina interdisciplinar y multidisciplinar, por lo que sus ámbitos de investigación son diversos y variados y han sido determinados por distintos autores.

Según María Teresa Turell (2005) estos ámbitos son:

1. El uso de información lingüística (tanto oral como escrita) para reconocer individuos de una determinada variedad lingüística, estilo o registro.
2. El análisis de la imitación de la firma y de la producción de textos con fines delictivos.
3. La atribución de autoría de textos orales y escritos.
4. El análisis del plagio.
5. La comprensión lectora de documentos legales.
6. El discurso legal y judicial.

¹ “We may say that Forensic Linguistics is the interface between language, crime and law, where *law* includes law enforcement, judicial matters, legislation, disputes or proceedings in law, and even disputes which only potentially involve some infraction of the law or some necessity to seek a legal remedy” (p. 12).

² “The forensic linguist may quote observations from research undertaken in fields as diverse as language and memory studies, Conversation Analysis, Discourse Analysis, theory of grammar, Cognitive Linguistics, Speech Act Theory, etc” (p. 13).

7. La interpretación y traducción legal, judicial y de apoyo en contextos multilingües.

J. Olsson (2004) también enumeró las mismas áreas de la lingüística forense e incluyó estas otras:

1. La determinación de la modalidad de producción del texto.
2. La transcripción de declaraciones verbales.
3. Los derechos lingüísticos del individuo procesado.
4. El análisis de las declaraciones hechas por testigos para determinar su índice de credibilidad.
5. El estatus de las pruebas documentales.
6. El lenguaje y el discurso del juzgado. Relación entre los participantes del juzgado y el lenguaje que usan: prejuicios, choque cultura, etc.

Por último, Coulthard, Johnson y Wright (2017, citados en Sánchez, 2017) distinguen principalmente cuatro áreas de trabajo del lingüista forense:

1. La fonética forense.
2. La atribución de autoría.
3. El plagio.
4. El papel del testigo como experto.

Siguiendo los modelos anteriores finalmente podemos concluir que los campos de actuación fundamentales de la lingüística forense son (Sánchez, 2017):

1. La fonética forense (elaboración de perfiles lingüísticos -pasaporte vocal-, reconocimiento de voces por parte de testigos, autenticación de grabaciones, determinación del contenido de grabaciones, reconocimiento de locutores).
2. La atribución de autoría (determinación de autoría, atribución de autoría y elaboración de perfiles lingüísticos).
3. El análisis del plagio.

Guiándonos por esta última clasificación de Sánchez (2017) explicaremos brevemente en qué consiste cada uno y en el siguiente capítulo conoceremos con mayor profundidad los principios de la atribución de autoría ya que nos será de gran utilidad para desarrollar nuestro estudio.

Primeramente, la fonética forense se encarga del análisis de la voz y el sonido en los casos criminales (Sánchez, 2017). Es el campo de actuación más desarrollado dentro de la lingüística forense, especialmente en España (Bernal, Riveiros, Garayzábal, 2014). Con la identificación de locutores la policía puede averiguar con un grado de certeza alto si el sospechoso ha emitido el mensaje o no (Bernal, Riveiros, Garayzábal, 2014). Los

investigadores recurren a la acústica forense cuando disponen de registros de audio como en los casos de amenazas telefónicas (Bernal, Riveiros, Garayzábal, 2014).

Para responder a preguntas como: ¿quién es el autor del texto escrito? O ¿es el sospechoso el autor del texto? los lingüistas forenses recurren a la atribución de autoría. Es decir, la atribución de autoría es de utilidad en casos en los que hay sospechas de que un texto determinado no ha sido escrito por el supuesto autor sino por otra persona (Bernal, Riveiros, Garayzábal, 2014).

Por último, plagiar consiste en copiar ilegalmente (y de manera total o parcial) obras o ideas sin reconocer al autor original de las mismas (Sánchez, 2017). La detección de plagio según Bernal, Riveiros y Garayzábal (2014) es una tarea controvertida pues, a pesar de que el grado de parecido de dos textos se exprese en porcentajes, los procedimientos pueden ser subjetivos. Por ello, los lingüistas forenses recurren a programas informáticos como Varbrul 2 y 3 y Goldvarb X para identificar si hay plagio o no e incluso llegar a conocer la autoría del texto (Bernal, Riveiros, Garayzábal, 2014).

3.3. Origen:

Como ocurre con casi todas las ciencias, no es posible localizar el nacimiento de la disciplina en un momento determinado. Lo que sí se ha podido especificar es el momento exacto en el que el término *forensic linguistics* fue empleado por primera vez. Fue en 1968, cuando Jan Svartvik, profesor de lingüística en la Universidad de Lund realizó un estudio lingüístico sobre las declaraciones del señor Timothy John Evans³, años después de que sucediera el crimen. Evans fue considerado como principal sospechoso del crimen que implicaba la muerte de su mujer e hija el 30 de noviembre de 1949. Tras sus cuatro declaraciones (dos en la comisaría de Merthyr Tydfil y otras dos en la de Notting Hill), fue condenado finalmente como culpable y ejecutado el 9 de marzo de 1950. En estas declaraciones (con grandes contradicciones entre sí) Evans se auto inculpó del crimen en dos ocasiones y en las dos restantes se declaraba inocente y señalaba como asesino a su vecino, el señor John Reginald Halliday Christie. Tres años más tarde, seis cuerpos se encontraron en la casa de Christie, entre ellos el de la esposa de Evans, confirmando entonces su culpabilidad y revelando que Christie era un auténtico asesino en serie. Finalmente fue condenado a muerte el 15 de julio de 1953. Posteriormente, en 1966 se dio el perdón póstumo a Timothy John Evans liberándole

³ Toda la información sobre el caso ha sido extraída de las obras de McMenamin (2002) y Ramírez Salado (2017) y del propio Svartvik (1968).

de toda culpa. A pesar de que la policía llegara a la verdad de los sucesos y descubriera el culpable de los asesinatos, este caso parecía no estar completamente cerrado pues no se conocía la razón por la que Timothy Evans había cambiado sus declaraciones. Es entonces cuando el lingüista sueco Jan Svartvik realizó su análisis lingüístico. En su estudio, Svartvik determina que (a pesar de que es difícil conseguir una conclusión estable), las declaraciones en las que Timothy inculpa a Christie son cuantificable y lingüísticamente distintas de las que se declara así mismo culpable porque estas últimas presentan características que no son propias del estilo gramatical de Evans. Gracias a este primer análisis y a la aportación de Svartvik a la investigación de tipo legal, se abrió una nueva puerta para los lingüistas. Aparecieron nuevos colectivos profesionales, asociaciones (IAFPA⁴ en 1991 y IAFL⁵ en 1992) y se fortaleció la figura del lingüista en la sociedad.

3.4. Situación en España y perspectivas actuales:

La lingüística forense en España es un área emergente de investigación a la que todavía no se le presta la suficiente atención debido a la falta de profesionales capacitados para realizar análisis de carácter lingüístico (Bernal, Riveiros, Garayzábal, 2014). Según María Teresa Turell (2005), en España y en otros países hispanohablantes no es común la colaboración entre lingüistas y juristas y, por tanto, es muy poco frecuente encontrar trabajos compartidos entre estos dos tipos de profesionales. Es por ello por lo que en España aún no se haya reconocido del todo la aportación de los lingüistas al campo judicial y policial a diferencia de países como EE. UU. y Canadá, que sí la valoran desde los años 60 (Bernal, Riveiros, Garayzábal, 2014).

En primer lugar, la lingüística forense tiene presencia académica en varios puntos de España. La Universidad Pompeu Fabra de Barcelona fue la pionera en lingüística forense, creando en 1993 un laboratorio de la mano de María Teresa Turell llamado ForensicLab en el que se realizaban peritajes lingüísticos (ForensicLab, <https://www.upf.edu/es/web/uval/forensiclab>). También se desarrollaban actividades formativas y de investigación en lingüística forense. Actualmente, en esta universidad, en la de Barcelona y en la de Nebrija se pueden estudiar grados en Lingüística, Lenguas Aplicadas y Lingüística Aplicada y Ciencias del Lenguaje respectivamente (Ramírez Salado, 2017). Respecto a otras universidades, tanto en la Universidad Complutense de Madrid como en la Universidad de Cádiz es posible estudiar el

⁴ International Association for Forensic Phonetics and Acoustics.

⁵ International Association of Forensic Linguists.

grado en Lingüística y Lenguas Aplicadas (aunque solo la Universidad de Cádiz recoge una asignatura específica sobre la lingüística forense). Además, en el Grado en Criminología y Seguridad de la misma universidad se ofrece la asignatura *Lingüística forense aplicada al delito*. Por otro lado, también es posible acceder a másteres que, aunque no ofrecen el estudio de la lingüística forense en profundidad, proporcionan herramientas útiles para el estudio de esta. En concreto, tanto la Universidad de Vigo como la de la Coruña (aparte de la Universidad Nacional de Educación a Distancia) tienen másteres relacionados con la Lingüística Aplicada (Ramírez Salado, 2017).

A pesar de que en España no existan asociaciones o grupos nacionales que promuevan y difundan actividades, congresos o estudios de lingüística forense (como la IAFL o IAFPA), en los últimos años han surgido grupos independientes interesados en nuestra materia (Ramírez Salado, 2017). En el ámbito privado, España cuenta con el laboratorio SQ-Lingüistas forenses, que realiza peritajes lingüísticos y además proporciona cursos de formación en lingüística forense (Ramírez Salado, 2017). Su directora, la Dra. Sheila Queralt es perito judicial, doctora en Traducción y Ciencias del Lenguaje, máster en Lingüística Forense (entre otros) y fue también investigadora del ForensicLab hasta 2017 (SQ-Lingüistas Forenses <https://www.sq-linguistasforenses.com/sheilaqueralt.html>). Su especialidad es la comparación forense de textos escritos y el análisis del discurso. Trabaja junto con la Dra. Núria Gavalda y la Dra. Montse Marquina para seguir con la labor de María Teresa Turell. El equipo ofrece un amplio número de servicios relacionados con la lingüística forense como son: la atribución de autoría, la identificación de la voz, la construcción de perfiles, el análisis del lenguaje criminal, la autenticación, limpieza y transcripción de grabaciones, la detección de plagio, la interpretación lingüística de ambigüedades y el análisis lingüístico de alteraciones, el análisis lingüístico de la identidad digital y de marcas comerciales, la evaluación de traducciones jurídicas y judiciales y el análisis del nivel de lengua de exámenes de lengua.

Por otro lado, España también cuenta con grupos académicos cuyos objetivos son difundir, formar e investigar sobre la lingüística forense. Entre ellos destacan el grupo ILFE, creado en 2009. Este grupo multidisciplinar organiza jornadas informativas de lingüística forense, dispone de cursos y talleres centrados en los diferentes ámbitos de esta disciplina y realiza informes periciales basados en la evidencia oral o escrita (Grupo Académico ILFE <https://grupoacademicoilfe.weebly.com/>). El equipo encabezado por la Dra. Elena Garayzábal Heinze cuenta también con la colaboración de la Dra. Sheila Queralt y Mercedes Reigosa Riveiros. Además, contamos también con un gabinete de lingüística forense que forma parte

del Instituto Universitario de Investigación en Lingüística Aplicada (ILA) de la Universidad de Cádiz cuyo director es el Dr. Miguel Casas Gómez (ILA <https://ila.uca.es/>). El ILA investiga sobre distintos ámbitos de la lingüística además de organizar cursos, talleres y seminarios de formación relacionados con la misma. Sus principales líneas de actuación son:

- La asesoría lingüística.
- La lingüística clínica.
- La lingüística forense.
- Las tecnologías del lenguaje e industrias de la lengua.

Por su parte, tanto la policía nacional como la guardia civil cuentan con laboratorios de acústica para resolver crímenes en los que es necesario un reconocimiento de voz, de grabaciones... etc. Según la comisaria general de policía científica, “la Acústica forense es uno de los más complejos entornos de investigación de la Policía Científica, debido fundamentalmente al carácter multidisciplinar de sus distintas aproximaciones de análisis, y a la necesidad de otorgar un alto y continuado nivel de formación a sus expertos” (Policía Científica https://www.policia.es/org_central/cientifica/servicios/tp_acustic_foren.html). Además, añaden que, aunque el empleo de la tecnología y las aplicaciones digitales es necesario para llevar a cabo los análisis, también lo es la participación de un equipo de profesionales y expertos especializados en distintas perspectivas de estudio. Las actividades que la policía nacional realiza para obtener indicios y pruebas que ayuden a la investigación son:

1. Estudios sobre identificación de locutores.
2. Estudios sobre manipulación de registros, procesado y edición de la señal de sonido.
3. Estudios de pasaporte vocal. Estos estudios parten de un registro hablado del que se establecen los rasgos de identidad (edad, sexo, etc.), las asociaciones diatópicas (área geográfica del hablante), las diastráticas (estrato social), emocionales, conductuales, patológicas o toxicológicas del hablante sospechoso o la víctima.
4. Identificación de fuentes de registro.
5. Ruedas de reconocimiento de voz.
6. Análisis y determinación de falsificaciones y pirateo de soportes magnéticos de audio, en colaboración con otras Unidades.
7. Estudios de registros no vocales (sonidos, ruidos de fondo, etc).

8. Acústica de disparos. Determinación de tipo de arma utilizada, ambiente acústico de la escena del crimen, etc.

Finalmente, el equipo de criminalística de la guardia civil, formado por investigadores forenses, técnicos y científicos, trabaja en la realización de informes periciales (con la ayuda de técnicas, equipos y procedimientos de investigación) con el fin de esclarecer delitos, sus culpables y sus víctimas (Guardia Civil <http://www.guardiacivil.es/es/institucional/Conocenos/especialidades/InvestigacionCientifica/index.html>).

4. La atribución de autoría. Conceptos generales:

Como hemos visto anteriormente, la lingüística forense es un campo interdisciplinar y multidisciplinar que dedica su estudio a diferentes ámbitos. Entre ellos se encuentra la atribución de autoría. En este capítulo definiremos este ámbito de actuación, sus conceptos generales y también los problemas que el lingüista forense se puede encontrar a la hora de identificar al autor de un texto. Explicaremos con mayor detalle en qué consiste pues nuestro estudio pretende ser un ejercicio de atribución de autoría.

El reconocimiento de emisores puede ayudar a que el investigador identifique al autor de un texto escrito cuando existen sospechosos que podrían o no haberlo escrito. Por tanto, es útil en los casos en los que un texto parece no haber sido escrito por su supuesto autor sino por otra persona (Bernal, Riveiros, Grayzábal, 2014). Es decir, la atribución de autoría forense es la tarea que se encarga de identificar al autor de un texto cuando el lenguaje es la evidencia que se usa en un juzgado (Olsson, Luchjenbroers, 2014, citados en Crespo, 2018). Este tipo de análisis estilístico puede identificar, describir y medir el estilo individual (o idiolecto) de un autor observando detalladamente el conjunto de sus elecciones lingüísticas (Guillén Nieto, 2008, citado en Crespo, 2018). Cada vez que un individuo escribe un texto, este será único, idiosincrático⁶ y estará compuesto por una serie de marcadores y recursos lingüísticos que lo harán característico (Turell, 2005). Pascual y Urbón (2009) también defendieron la idea de que un autor, al escribir, elabora un discurso que (a no ser que lo configure intencionalmente para ocultar su identidad) se caracteriza por unas marcas verbales y unos patrones de escritura

⁶ Según Picornell (2014) “las idiosincrasias son rasgos léxicos, sintácticos y de personalidad que tienden a constituir los elementos más distintivos de la producción textual de un autor en particular” (p. 83). Es decir, las idiosincrasias son un tipo de variación lingüística, son las formas específicas que escoge un autor cuando escribe (McMenamin, 2002).

que lo hacen diferente de los discursos de otras personas. Para estos autores, cada escritor posee un ideolecto⁷ propio y exclusivo.

El análisis de autoría se divide en dos tipos de procedimientos dependiendo del objetivo que persiga el investigador (Picornell, 2014):

1. La determinación del perfil lingüístico, en la que el lingüista forense elabora el perfil del individuo tras el análisis de los rasgos que caracterizan a sus textos y de los que se puede conocer: el género, la edad, el nivel educativo... etc. Este tipo de análisis suele ser empleado cuando en una investigación no existen sospechosos.
2. La atribución de autoría, por la que se comparan los rasgos lingüísticos de dos o más textos escritos por uno o más autores, de autoría conocida o anónima. Con este análisis se puede averiguar si los textos son de un único autor o de varios o incluso descubrir el autor de los textos.

La elaboración de perfiles lingüísticos es un estudio útil en casos en los que no se ha identificado ningún sospechoso y, por tanto, la policía necesita tener idea sobre las características sociales del escritor del texto (Crespo, 2018). Este autor también afirma que asociando los rasgos lingüísticos de los autores con grupos sociales y geográficos se logrará estrechar el rango de los posibles sospechosos del crimen que es el objetivo fundamental de la elaboración de perfiles.

Según Queralt (2014) el lingüista forense debe tener en cuenta estas dos reglas para decidir si es posible la identificación de un perfil lingüístico o no:

- 1) La muestra debe ser suficientemente larga.

Si el texto no es lo suficientemente amplio no será posible encontrar sus rasgos identificativos. Desafortunadamente, los textos con los que el lingüista forense trabaja normalmente (notas de rescate, notas de suicidio, amenazas...) suelen ser bastante cortos como para poder hacer una identificación fidedigna (Olsson y Luchjenbroers 2014, citados en Crespo, 2018). Para que el análisis de autoría sea fiable, se debe comparar un número suficiente de textos tanto de la víctima como del sospechoso y así poder identificar las características lingüísticas de los autores (Bernal, Riveiros, Garayzábal, 2014). El problema está, opina Fitzgerald (2014), en que cualquier persona puede cambiar sus conductas lingüísticas con el fin de ocultar su identidad o hacerse pasar por otra persona. Para utilizar el texto con fines delictivos y confundir al analista, el escritor podrá utilizar un pseudónimo, añadir errores ortográficos a propósito... etc. Es por ello por lo que es de gran importancia

⁷ Para Pascual y Urbón (2009) el ideolecto son unos hábitos personales del escritor que escapan a lo consciente.

que el lingüista forense disponga de una cantidad de textos considerable pues sus resultados serán más exactos (Fitzgerald, 2014).

- 2) La muestra debe ser de calidad. Para que sea de calidad, el texto deberá reflejar el estilo de su escritor, es decir, marcas de su dialecto, sexo, nivel de estudios... etc.

La selección de los rasgos característicos y relevantes de un autor no es fácil. Es cierto que cada persona usa el lenguaje de manera diferente, pero esto no significa que su estilo lingüístico se asemeje a una “huella dactilar” y por tanto su identificación no es de fácil acceso (Olsson, 2004, citado en Crespo, 2018). Según Picornell (2014) se pueden estudiar más de 1000 posibles marcadores. A pesar de que, no exista una lista de los marcadores lingüísticos útiles para analizar el perfil lingüístico de un autor, los investigadores forenses suelen seleccionar los rasgos con mayor frecuencia de aparición, los que no se puedan eliminar voluntariamente y los que sean comunes a la mayor parte de los individuos de la población de estudio (Picornell, 2014). Según Stefanova Spassova (2009) las marcas que el investigador debe seleccionar para identificar a un escritor deben tener las siguientes características:

1. Prominencia. Es decir, deben destacar estadísticamente cuando se compara un corpus con otro.
2. Independencia relativa del control consciente, que tiene que ver con el grado en el que el escritor es capaz de controlar sus elecciones lingüísticas cuando produce un texto. Cuanto menos manipulable sea la marca que se quiere estudiar, más adecuada será para que sea una marca identificativa.
3. Distribución y uso frecuente.
4. Cuantificación fácil.

Por otra parte, y a modo de conclusión de este capítulo, McMenamin en su libro *Forensic Linguistics, Advances in Forensic Stylistics* (2002) nos da la clave para entender por qué la pragmática puede ser de gran utilidad para la lingüística forense. Según el autor, la pragmática es el análisis del significado intencional de un hablante en contextos de uso reales. A su vez, la lingüística forense, usa las herramientas teóricas de la pragmática para analizar la función del lenguaje usado en esos determinados contextos. Por todo esto, McMenamin opina que “pragmatics is important for forensic purposes because speakers and writers do not always directly match their words with the meaning that they intend to convey. Since listeners and readers may also be unsuccessful in matching expression to intended meaning, the speaker’s or writer’s intended meaning is more open to interpretation by the listener or reader, sometimes resulting in mistaken understanding, miscommunication, and, eventually, conflict”

(p. 93). Por su parte, Isabel Picornell (2014) también consideró que la pragmática juega un papel fundamental para la lingüística forense pues para la atribución de autoría es necesario comparar textos con un contenido similar y escritos en el mismo nivel de (in)formalidad.

5. El registro coloquial como instrumento de análisis forense español:

5.1. El registro coloquial en español.

5.1.1. El registro coloquial y la conversación.

5.2. Marcas identificativas del registro coloquial.

5.2.1. Marcas morfosintácticas.

5.2.2. Marcadores pragmático-discursivos.

5.1. El registro coloquial en español:

Desde hace tiempo se ha insistido en la necesidad de desarrollar una lingüística del hablar, esa gramática del hablar, del estudio del texto en su contexto (Briz, 2001). “Ya en los años sesenta el estudio de estas relaciones se convierte en objetivo de lingüistas, semióticos, filósofos, psicólogos, antropólogos y sociólogos. Y cada vez más en la actualidad se insiste en la prioridad, aunque no exclusividad, del estudio del funcionamiento oral de la lengua y de las formas dialogadas” (Briz, 2001, p. 9).

En el marco de esa necesidad general, este capítulo dirará en torno al concepto de registro como elemento indispensable a la hora de codificar el mensaje en una situación comunicativa determinada. Veremos qué tipos de registros pueden identificarse y hablaremos sobre el registro coloquial en español (en qué consiste, cómo se caracteriza y en qué se diferencia del registro formal).

A pesar de no ser el primero en hablar sobre el español coloquial, Antonio Briz es uno de los autores más influyentes que desarrolla sus principios. De ahí que debamos comenzar este apartado desarrollando de forma general sus aportaciones sobre el tema.

Según Briz (2001), la lengua cambia dependiendo de cuatro factores: el tiempo (variedad diacrónica), el espacio (variedad diatópica), las características de los usuarios (variedad diastrática) y la situación de comunicación (variedad diafásica). Estos efectos, producen estados sincrónicos diferentes a lo largo de la historia de la lengua, dialectos, sociolectos y registros (Briz, 2001). Los registros son las modalidades de uso que dependen del contexto comunicativo (Briz, 2001).

Debido al contexto comunicativo, se regulan las conductas lingüísticas y extralingüísticas de los hablantes ya que estos deben adaptar en mayor o menor grado sus actos lingüísticos a la situación determinada en que tienen lugar (Briz, 2001), ya que, si el hablante no adapta sus usos lingüísticos a la situación dada, provocaría desajustes no informativos, pero de conducta lingüística convencional o esperable (Briz, 2001). De esta manera, las expectativas del destinatario se romperían si el registro esperado no se usa en la situación adecuada.⁸

Para Briz, el dominio y el buen empleo de estos registros lingüísticos está relacionado con el nivel de lengua de los usuarios. Se trata de una relación proporcional por lo que, a mayor nivel de lengua, mayor dominio de registros.

Existen dos tipos de registros: el formal y el informal o coloquial. La relación de proximidad entre los participantes, su conocimiento del mundo y experiencias compartidas, la cotidianidad, el grado de planificación y la finalidad de la comunicación son las características que, para este autor, hacen posible la distinción entre un tipo de registro u otro. Cuando afirmamos que del uso que un hablante en una situación determinada hace de una lengua resultan los registros, aceptamos también que en el empleo de los registros existe una correlación entre los rasgos situacionales y las propias características del hablante (que dotarán al registro de matices dialectales y sociolectales). Esto puede traducirse en que, por ejemplo, el español coloquial de un hablante andaluz será diferente al de un hablante valenciano o que, por otro lado, el español coloquial de un hablante que pertenece a un nivel sociocultural alto presentará diferencias del español coloquial de un hablante de nivel sociocultural más bajo. No se trata, en suma, de un rasgo unificado y con las mismas características en todos los perfiles de hablantes, aunque sí parece posible, como detallaremos más adelante, trazar una serie de rasgos generales susceptibles de aparecer o no representados en el registro coloquial que desarrolle un determinado perfil de hablantes.

Además, debemos tener en cuenta que cada registro está vinculado a modos de comunicación diferentes (Givón, 1979 y Ochs, 1979, citados en Briz, 1998). Existen dos modos diferentes de comunicación: el modo pragmático y el modo sintáctico. El modo pragmático se adquiere de forma progresiva y natural por medio de la transmisión, del contacto lingüístico con otros hablantes, es derivado de un proceso de adquisición (Briz, 1998). Por otro lado, el modo sintáctico se aprende mediante un proceso de enseñanza, generalmente iniciado en la escuela, es el resultado de un proceso de aprendizaje. El modo pragmático se diferencia del modo sintáctico en que este último es un aprendizaje consciente, mientras que el pragmático

⁸ Así se explotaría la convención, según Paul Grice.

consiste en adquirir formas más naturales y familiares (Briz, 1998), si bien ambos modos se actualizan de acuerdo con la situación de comunicación (Briz, 1998). Del desarrollo del modo pragmático resulta el uso coloquial de la lengua, y del modo sintáctico se deriva el uso formal de la lengua⁹.

Todo usuario de la lengua que domine estos modos de expresión será capaz de usarlos adecuadamente según el contexto. Esto es precisamente la pragmática y no solo el empleo de la coloquialidad.

El uso de los registros, formal y coloquial, puede manifestarse oralmente o por escrito, a pesar de que en esta última forma siempre existe un mayor grado de formalidad (Briz, 2001). Por tanto, Briz (2001) establece al menos cuatro realizaciones discursivas: coloquial oral, coloquial escrito, formal oral y formal escrito.

Briz (2001) reconoce que, aunque la modalidad de uso coloquial es más frecuente en el habla que en lo escrito (a diferencia de la modalidad formal), asociar el registro formal con lo escrito (lo literario) y relegar el registro coloquial a lo meramente oral sería un error. El autor llega a esta idea tras presentar un ejemplo de una carta familiar en la que el emisor combina una estructura discursiva que posee características propias de una carta (como la fecha, la petición de perdón por tardar tanto en escribir, la despedida, la firma, la posdata...) con un claro tono informal y cotidiano expresado en el léxico y la fraseología utilizada. Por lo que, aunque se haya asociado con frecuencia el registro formal con lo escrito, la oralidad coloquial y el modo de escritura pueden aparecer en una misma producción. (Briz, 2001). Además, los usuarios de una lengua que solo tienen conocimiento de la variedad coloquial son dados a reflejarla en sus producciones escritas de forma natural (Briz, 2001) (es decir, escriben tal y como hablan). En estas producciones, aparecen la acumulación de enunciados, los anacolutos, las ambigüedades e incluso la propia reproducción de la oralidad. Un ejemplo de ello sería la siguiente expresión: “Proivido zalta la lan bra”¹⁰ Este comportamiento también parece manifestarse hoy en día en el ámbito de las nuevas tecnologías, sobre todo en el creciente uso de las redes sociales y mensajería instantánea (Twitter, Facebook, Whatsapp...) en las que sus

⁹ Cabe cuestionar esta correlación entre lo pragmático y lo coloquial por un lado y lo sintáctico y lo formal por otro. Relacionar el ámbito de la pragmática únicamente con las manifestaciones en registro coloquial puede resultar aventurado, en la medida en que la pragmática estudia el empleo del código en relación con los factores de la situación comunicativa concreta en la que este se emplea, independientemente del registro utilizado por el emisor. M. V. Escandell (2003, p. 14) definió la pragmática como “una disciplina que toma en consideración los factores extralingüísticos que determinan el uso del lenguaje, precisamente todos aquellos factores a los que no puede hacer referencia un estudio puramente gramatical: nociones como las de emisor, destinatario, intención comunicativa, contexto verbal, situación o conocimiento del mundo que van a resultar de capital importancia”.

¹⁰ Artículo L. Carandell publicado en el diario El País. Letrero encontrado sobre la alambrada que rodeaba una finca en Granada.

usuarios son propensos a escribir todo tipo de mensajes e ideas de manera rápida, informal y con una recepción inmediata. Volveremos sobre esta cuestión en la parte práctica de nuestro trabajo.

En definitiva, el autor (Briz, 2001) concluye que un usuario de una lengua, al hablar y escribir (según la situación de comunicación) puede usar el registro coloquial (conversación informal entre amigos, carta familiar), el registro formal (conferencia en un salón de actos, texto literario) o imitar uno de los registros (queja elevada al ayuntamiento por parte de un hablante perteneciente a un estrato sociocultural bajo).

Centrándonos ahora en el registro coloquial, pasaremos a analizar las diferentes definiciones propuestas por diversos autores. El propio Briz (2001) lo ha denominado de varias maneras: *español coloquial*, *habla coloquial*, *registro coloquial*, *uso coloquial* o *modalidad lingüística coloquial*. Incluso tampoco rechaza usar los términos *lengua o lenguaje coloquial* (Briz, 2001). Sí se rechaza sustituir el término *coloquial* por *conversacional* pues de esta manera se puede llegar a confundir un tipo de discurso con el registro informal¹¹.

Finalmente, usar *vulgar* y *popular* como sinónimos de coloquial sería inadecuado. Para Briz, el término *vulgar* alude al uso incorrecto o al margen de la norma estándar, por lo que no es posible intercambiarlo por *coloquial* porque el registro coloquial es “un uso socialmente aceptado en situaciones cotidianas, no vinculado en exclusiva a un nivel de lengua determinado y en el que vulgarismos y dialectalismos aparecen en función de las características de los usuarios” (1996, citado en Briz, 2001, p. 25-26). Por otro lado, según M. Seco (1973, citado en Briz, 2001) con el término *popular* se alude a las características socioculturales del hablante y esto pertenece a un nivel de lengua y no de habla (el registro coloquial pertenece a este campo).

Pero el problema no está en cómo denominar este fenómeno sino más bien en llegar a definirlo, a enumerar sus características principales.

No es fácil definir el término pues, intuitivamente, podemos llegar a asociar una serie de características al registro coloquial que no se corresponden con lo que verdaderamente es dicho registro (Briz, 2001). Para Seco (1973, citado en López, 2007) el problema de la definición del término “registro coloquial” se encuentra en identificar la lengua coloquial con la lengua popular informal sin tener en cuenta la vertiente coloquial en la lengua “media”.

¹¹ Otra forma de denominar el registro informal sería usando el término *familiar* (término usado por J. Polo en sus trabajos 1971-1976, citado en Briz, 2001) siempre que se entienda como hecho cotidiano y no como “relación entre personas allegadas” pues personas que no tienen una relación cercana pueden hablar coloquialmente.

Para Antonio Narbona (2015) las notables discrepancias al jerarquizar las numerosas nociones a las que se ha recurrido para definir y caracterizar el español coloquial derivan del hecho de no poder integrar la *variación* y las *variedades* en la explicación.

Este tipo de registro se relaciona con lo cotidiano y lo informal y tiene carácter espontáneo y natural. Además, es expresivo y subjetivo y se caracteriza por su falta de planificación. Sin embargo, no se debe confundir lo coloquial con lo vulgar. Por otro lado, es importante diferenciar entre registro y sociolecto (entre nivel de habla y de lengua). Otro de los errores más comunes, como ya hemos comentado antes, es identificar el registro coloquial como exclusivamente oral y relegarlo al español hablado.

Beinhauer (1991, citado en Briz, 2001, p. 37) definió el español coloquial como: “El habla tal como brota, natural y espontáneamente en la conversación diaria, a diferencia de las manifestaciones lingüísticas conscientemente formuladas, y por tanto más cerebrales, de oradores, predicadores, abogados, conferenciantes, etc., o las artísticamente moldeadas y engalanadas de escritores, periodistas o poetas”¹². Para Beinhauer, los rasgos del registro coloquial “no constan tan solo de elementos sintáctico-estilísticos por un lado, y de vocablos y giros, o sea de elementos lexicológicos, por el otro; a todos ellos se agregan los medios dinámicos de entonación, gesto y mímica” (López, 2007, p. 5-6). Asimismo, este autor destaca que es un error muy común confundir el lenguaje cotidiano hablado con la lengua cotidiana escrita o impresa (Beinhauer, 1978).

También, M. C. Lasalletta definió el español coloquial en 1974 como “[...] una fracción o nivel de lenguaje total que se destaca por su carácter pintoresco reflejado en multitud de expresiones y vocablos intraducibles a otros idiomas, fundados muchas veces en alusiones metafóricas y que posee una gracia, viveza, gran espontaneidad, concreción y expresividad que lo distinguen de otros niveles [...]” (Briz, 2001, p. 37-38). Otra definición destacable fue aportada por B. Steel en 1976: “The term colloquial [...] is commonly felt -albeit often pejoratively- to refer to particular informal (often racy or popular) spoken usage, especially that usage which differs in some way from formal language [...]” (Briz, 2001, p. 38).

Por su parte, Julio Casares define el registro como ‘lengua de la conversación’ o ‘estilo familiar’ (López, 2007). La definición de Tudora Sandru Olteanu (1988, citado en López, 2007, p. 11) afirma que “el lenguaje coloquial [...] es la modalidad lingüística empleada natural y espontáneamente por los hablantes en sus relaciones informales de la vida cotidiana”.

¹² Con “lenguaje coloquial” el autor solo se refiere a la lengua viva conversacional (Beinhauer, 1978).

En el reconocimiento del registro coloquial destacan los estudios de E. Lorenzo (1977) que describen y analizan los elementos que configuran el coloquio (según Val (1980) “la suma elemental de dos o más interlocutores con significado complementario” (Briz, 2001, p.38)). Los elementos son: la situación, los contextos, el tipo de mensaje, la confrontación de fuerzas entre los interlocutores o tensión coloquial, los efectos de la voz y las pausas (Briz, 2001). E. Lorenzo define la lengua española coloquial como: “El conjunto de usos lingüísticos registrables entre dos o más hispanohablantes, conscientes de la competencia de su interlocutor o interlocutores, en una situación normal de la vida cotidiana, con utilización de los recursos paralingüísticos o extralingüísticos, aceptados y entendidos, pero no necesariamente compartidos, por la comunidad en que se producen” (Briz, 2001, p. 38). El mismo autor hace una distinción entre las constantes y las variables del registro coloquial español. Entre las constantes, se encuentran la presencia física de uno o más hablantes y el marco espaciotemporal que sirve de referencia a la comunicación. Las variables son la experiencia común (que depende del grado de convivencia que exista entre los hablantes), su carácter déictico, su egocentrismo, etc.

Sin embargo, este tipo de estudio y definición del registro coloquial puede llegar a generar cierta confusión entre un tipo de discurso (la conversación) y una modalidad de uso (coloquial) en la que la conversación se puede usar (Briz, 2001). Esto mismo ocurre en la definición de A. M. Vigara (1992, citada en Briz, 2001, p. 39) “lo que designamos con el sintagma español coloquial es el empleo común que hacen de un determinado sistema lingüístico los hablantes de una determinada sociedad (la española) en sus actos cotidianos de comunicación. [...] La conversación (o coloquio) no es, en suma, sino una forma de interacción verbal puntual, determinada por tres características que le son consustanciales: la *actualización oral*, su *inmediatez* y la *interdependencia dinámica de todos los elementos* en el proceso de la comunicación”.

M. Josefina Tejera (1989, citada en López, 2007, p. 17) reconoce que los lingüistas están de acuerdo en que “para cada situación de comunicación existen variedades funcionales contextuales del lenguaje, que son los llamados estilos, tonos, estratos, niveles o registros de lengua, los cuales establecen las diferencias diafásicas [...]”. En lo que no se ponen de acuerdo es en los límites entre los niveles ni en su nomenclatura (López, 2007).

Conociendo todas estas definiciones, Briz (2001) resume que el español coloquial:

- Es un registro y un uso que emplea el hablante de acuerdo con la situación y a las circunstancias de la comunicación en la que se encuentre.

- No pertenece a ninguna clase social en concreto, sino que singulariza las realizaciones de los hablantes de una lengua. Aunque sea un registro dominado por los individuos de nivel sociocultural bajo o medio, en absoluto les es exclusivo¹³.
- No es uniforme ni homogéneo. Cambia según las características dialectales y sociolectales de los hablantes. Debemos matizar que, como comprobaremos más adelante en nuestro estudio, esta afirmación debe cuestionarse.
- Refleja un sistema de expresión que antes que ser un uso simplificado del registro formal, es la continuación y el desarrollo del *modo pragmático* de la comunicación humana.
- Es un registro que se manifiesta tanto de modo oral como de modo escrito.
- Aparece en distintos tipos de discurso, aunque en la conversación es donde el registro coloquial se manifiesta de manera más natural.

Según los trabajos de Halliday, McIntosh y Stevens (1964, citados en Briz, 2001) y Gregory y Carroll (1978, citados en Briz, 2001) los criterios que definen los registros son: el *campo*, el *modo*, el *tenor* y el *tono*. El campo se refiere al carácter técnico o no del discurso (hechos cotidianos, científico, legal...); el modo se limita al canal o medio (escrito o literario, hablado espontáneo...); el tenor es el papel que el lenguaje juega en esa situación comunicativa y el tono queda limitado a las relaciones de formalidad entre los hablantes y al grado de formalidad en la comunicación.

De este modo, el registro coloquial quedaría delimitado así (Payrató, 1992, citado en Briz, 2001, p. 41):

CAMPO: cotidianidad

MODO: oral espontáneo

TENOR: interactivo

TONO: informal

Aunque Briz aprobara la delimitación del registro coloquial que acabamos de mencionar (como un primer acercamiento), consideró que como la situación es un factor fundamental en el empleo del registro coloquial, para caracterizarlo y conocer su contexto de uso es conveniente distinguir también los rasgos asociados a la situación que favorecen su uso

¹³ Cabe cuestionar si procede asociar el dominio de un determinado registro a un estrato sociocultural dado ya que el propio Antonio Briz estableció tres parámetros fundamentales en relación con el registro: el lugar, el tema y la relación entre los hablantes y nunca el estrato sociocultural.

(rasgos situacionales) de otros rasgos primarios que se intuyen tras un uso determinado del lenguaje en esa situación.

Los rasgos situacionales o colonizadores son:

- La *relación de igualdad* entre los interlocutores, social (nivel sociocultural, profesión) o funcional (el papel que poseen en una situación determinada). Si los hablantes se relacionan como iguales es más frecuente la coloquialidad (Briz, 2001).
- La *relación vivencial de proximidad*, es decir, el conocimiento del mundo compartido y experiencias comunes.
- El *marco discursivo familiar* que depende del espacio físico y la relación que mantienen los participantes con ese espacio.
- La *temática no especializada*, la cotidianidad, los temas que están al alcance de cualquier hablante.

Cuanta más afinidad se dé entre los interlocutores con respecto a estos rasgos, mayor será la coloquialidad.

Por otro lado, los rasgos primarios (relacionados con los anteriores) que caracterizan el registro coloquial son:

- La *ausencia de planificación* o la planificación sobre la marcha que favorece la espontaneidad.
- La *finalidad interpersonal*, el fin comunicativo de socialización.
- El *tono informal* como resultado de todos los rasgos anteriores.

Los integrantes del grupo Val.es.co (Valencia, Español, Coloquial) dirigidos por Briz reconocen que aunque el uso los rasgos antes mencionados no definen la conversación coloquial, sí que sirven para reconocer en ella la modalidad de uso empleada y su grado de coloquialidad (Narbona, 2015).

5.1.1 El registro coloquial y la conversación:

El registro coloquial está, pues, estrechamente relacionado con la conversación. Para A. Briz, la conversación es un tipo de discurso que se caracteriza por ser una interlocución *presencial* cara-a-cara; *inmediata* (aquí y ahora); con *toma de turnos no fijada*; *dinámica* (la alternancia de turnos es sucesiva y simultánea) y *cooperativa*.

El hecho de que la conversación sea inmediata, la diferencia de la simple sucesión de mensajes; su toma de turnos no predeterminada la opone de otro tipo de discursos como la entrevista; y su carácter dinámico la distingue del discurso político o el intercambio de

saludos. La conversación será coloquial cuando presente los rasgos primarios que hemos mencionado anteriormente (Briz, 2001). Diferenciaremos entre conversaciones coloquiales prototípicas o periféricas según la mayor o menor presencia de los rasgos situacionales. Si la conversación coloquial se da entre dos amigos que hablan en la calle sobre lo que han hecho el fin de semana, la conversación es prototípica. Por tanto, toda “conversación no preparada, con fines interpersonales, informal, que tiene lugar en un marco de interacción familiar, entre iguales (sociales o funcionales) que comparten experiencias comunes y en la que se habla de temas cotidianos” (Briz, 2001, p.43) será considerada coloquial prototípica. Si en la conversación coloquial falla alguno de estos rasgos, pasará a considerarse como periférica (Briz, 2001).

Para fijar los principios que explican el desarrollo de la conversación como conducta social o negociación, P. Grice en 1975 y 1978 propuso el *principio de cooperación* y otros como H. Lakoff (1973), P. Brown y S. Levinson (1978, 1987)¹⁴, G. Leech (1983) y H. Haverkate (1994) desarrollaron el llamado *principio de cortesía*. Todas estas aproximaciones nos permiten explicar el uso coloquial conversacional como aparente “transgresión” (en ocasiones) de algunos de estos principios.

Cuando se mantiene una conversación también se interactúa y se negocia. Se produce una argumentación de ideas con el fin de conseguir un acuerdo con el otro. Conversar es comunicar cooperativamente y negociar de manera estratégica un objetivo con el otro. Por ello, en la conversación se persuade y se busca constantemente ser aceptado por el destinatario. Según Narbona (2015) conversar es persuadir y tratar de imponer los presupuestos propios modificando los del otro. Normalmente el emisor además de configurar su discurso en función a sus propios intereses también tiene en cuenta al oyente, al que incluso guía en sus movimientos. Para explicar esto, Briz hace alusión a S. Bonilla (1990) y a su empleo de la metáfora del lenguaje como juego elaborada por Wittgenstein. Según la metáfora, los jugadores (es decir, los hablantes) no solo respetan cooperativamente las reglas del juego, sino que también juegan para ganar (Briz, 2001). Por ello, el buen jugador de ajedrez mueve sus piezas teniendo en mente los movimientos que se han hecho antes y los que probablemente realizará su oponente. Esto mismo ocurre en la conversación, ya que cada enunciado e intervención de un hablante depende de los enunciados e intervenciones producidas anteriormente tanto suyas como del otro interlocutor y se prevén sus posibles

¹⁴ Aunque A. Briz los menciona como autores relacionados con el principio de cortesía, obvia la contribución fundamental de estos autores para el análisis de la conversación como instrumento para la negociación: el concepto de imagen.

respuestas. Cada intervención que se da en la comunicación es una restricción para la siguiente (al igual que en el ajedrez) (Bonilla, 1990, citado en Briz, 2001). Es decir, una intervención de inicio (por ejemplo, una pregunta) espera una de carácter reactivo (una respuesta) (Bonilla, 1990, citado en Briz, 2001).

Asimismo, el oyente colabora y actúa mostrando acuerdo o desacuerdo con lo dicho anteriormente. Además, todos sus movimientos futuros se sucederán hasta que se dé por finalizada la negociación. Por tanto, la conversación no avanzará si no existe cooperación entre los interlocutores. Según Grice, este principio de cooperación (por el cual, emisor y receptor destinatario presentan una actitud positiva previa a la comunicación) se regula mediante unas reglas. Las reglas o máximas conversacionales son: la máxima de *cantidad* (la información transmitida no debe ser excesiva pero tampoco insuficiente), la máxima de *cualidad* (la información debe ser verdadera y estar fundamentada con pruebas), la máxima de *relación* (la información es pertinente en relación con la situación comunicativa) y la máxima de *manera* (el emisor debe ser claro en su formulación). Sin embargo, cuando las intenciones de cada interlocutor no coinciden y para conseguir los efectos comunicativos deseados deben incumplirse estas reglas (Payrató, 1996, citado en Briz, 2001).

Los motivos por los que la atención al lenguaje conversacional es cada vez mayor son lo que nos hacen entender todas las dificultades con que se encuentra su desarrollo y análisis (Narbona, 2015). Este interés por la conversación deriva de la propia trayectoria de la lingüística, la cual se exigía una nueva forma de proceder y una nueva concepción del lenguaje. Su carácter marcadamente filológico se debía a que la escritura fue la que permitió el desarrollo de la conciencia metalingüística científica. Para dar el salto a lo oral no solo se necesitaban unas nuevas condiciones técnicas como la posibilidad de grabar y reproducir las actuaciones habladas, sino que también era necesario que se eliminara el prejuicio de considerar la lengua coloquial como deficitaria (calificada así tanto por sus carencias como por la creencia de que resultaba fallida) e inferior a la lengua escrita. Este prejuicio supuso el impedimento de abordar su examen de manera no mediatizada. Las anomalías o carencias que puedan presentar las producciones de hablantes con escasa competencia comunicativa no deben ser el foco de atención del lingüista. De hecho, para poder estudiar la técnica constructiva del habla, se debe dejar de ver sus usos como deficiencias derivadas de la incompetencia o torpeza del usuario y, por el contrario, optar por una nueva perspectiva que parta del proceso de enunciación y del modo de producción-recepción de las actuaciones propias de la inmediatez.

Por otro lado, como hemos plasmado durante el capítulo, el objeto de estudio presenta dificultades como la cantidad de términos empleados en alusión a un mismo fenómeno. Además, los problemas teórico-metodológicos todavía no están resueltos y tampoco hay un acuerdo sobre las unidades con que se debe operar.

Narbona afirma que intentar acotar una variedad del español, en este caso, la coloquial siempre es arriesgado, pues no hay que olvidar que se captan “solo aspectos parciales del evento comunicativo” (Oesterreicher, 2005, citado en Narbona, 2015). Además, cuantos más usuarios empleen una determinada modalidad de uso, más cautelosos debemos ser (Narbona, 2015).

“La conversación es claramente la manera más prototípica de usar el lenguaje” (Levinson, 1983, citado en Briz, 2001) y para Narbona (2015), la modalidad coloquial es el tipo de interacción comunicativa universal por excelencia, aunque no todos los usuarios de ella hablan coloquialmente de igual manera.

En definitiva, el registro coloquial es un registro reconocido en el ámbito de análisis léxico, si bien adolece, como hemos podido comprobar, de imprecisiones terminológicas y conceptuales que pueden llegar a dificultar la debida acotación del fenómeno. El alto grado de productividad comunicativa que en la actualidad va asociado con este registro (recordemos que su empleo no se ciñe solamente a la oralidad, sino que también se ha extendido ya a numerosos entornos comunicativos escritos vinculados con las tecnologías) es el que nos ha llevado a elegirlo como escenario para la construcción de nuestro análisis¹⁵.

¹⁵ Por otro lado, podemos afirmar que los estudios sobre el español coloquial son necesarios para la lingüística e incluso para el estudio de la literatura pues, como dijo Beinhauer (1978, p. 10) “quien niegue la [...] importancia de dicha materia [...] olvida que la lengua -incluso de poetas y literatos y aun más de eruditos, sobre todo los de habla española- arraiga profundamente en el subsuelo del lenguaje familiar y popular, del que se nutre a diario. Por tanto, sólo será capaz de sentir, captar y apreciar las ultimas intenciones [...] de un lenguaje artístico, quien conozca también la materia prima de que este está amasado, o sea, la lengua del pueblo, del ambiente en que vive el artista, la que este mismo habla a diario. Es más: no me recato en afirmar que quien no está debidamente familiarizado con el lenguaje coloquial, tampoco puede dominar realmente la lengua escrita. Podrá si acaso, a fuerza de estudiar gramática, llegar a expresarse con alguna corrección, pero esto no equivale, ni muchísimo menos, a lo que yo entiendo por dominio verdadero del idioma”.

5. 2. Marcas identificativas del registro coloquial:

En suma y, como hemos podido comprobar en el apartado anterior, el registro coloquial constituye una evidencia y un campo de estudio reconocido por la comunidad científica. Es un ámbito de estudio dotado de producciones lingüísticas específicas y rasgos que resultan identificables con nitidez. Este es el motivo fundamental por el que hemos decidido utilizarlo como herramienta para el desarrollo de la fase práctica de nuestro trabajo. En definitiva, utilizaremos algunos de los rasgos que la comunidad científica vincula con el empleo del registro coloquial en español para tratar de localizarlos en producciones concretas de varios perfiles de hablantes. Justificaremos nuestra selección de los citados rasgos retomando los dos modos de comunicación humana ya mencionados anteriormente: el *modo pragmático*, que el hablante adquiere de manera natural y mediante transmisión y contacto con otros hablantes, y el *modo sintáctico*, aprendido por medio de la enseñanza en la escuela.

En lo que supone una modificación parcial de los planteamientos de A. Briz, seleccionaremos expresiones coloquiales vinculadas con ambos modos, entendidos como escenarios distintos en los que se generan expresiones que diferentes autores relacionan con el uso del registro coloquial¹⁶.

Tras ello, verificaremos su grado de uso en wasap¹⁷, un canal nacido del desarrollo tecnológico aplicado a la comunicación humana entre hablantes de diferentes características. Nuestro objetivo es, por tanto, comprobar si es posible trazar perfiles de hablantes en lo que respecta al empleo de los rasgos elegidos en interacciones prototípicas de wasap para su uso en el reconocimiento de textos dubitados en el ámbito de la lingüística forense.

5.2.1. Marcas morfosintácticas:

En el nivel morfosintáctico las marcas elegidas son:

1. La coordinación inespecificativa: Y, PERO.

Ana María Vígara (2005) defiende que cuando el hablante emplea la coordinación lo hace con el fin de no explicitar la relación que une las partes del enunciado (p. 123). El uso del nexo inespecificativo y cuando el hablante construye una relación de contraste, hace que el receptor

¹⁶ Recordemos que Briz (2010) solamente relacionaba el empleo del registro coloquial con el modo pragmático y vinculaba el registro formal con el modo sintáctico.

¹⁷ El libro de estilo de la RAE recomienda escribirlo así y no con la forma inglesa.

deba descifrar dicha relación (p. 124). Este hecho hace que se potencie el contraste entre los miembros de la relación (Narbona y Morillo Velarde, 1987, citados en Vigara, 2005).

Además, el nexo *pero* también sirve como conector inespecificativo de enunciados. Según la autora, este tipo de coordinación en realidad es aparente pues, aunque se usen nexos coordinantes, la relación coordinativa de las partes no existe ya que ni el nexo *y* ni el nexo *pero* mantienen su verdadero valor en estos enunciados. Esto se debe a que incumplen los requisitos (exceptuando el de la entonación) que Bally (*Linguistique générale et linguistique française*, 1932, citado en Vigara, 2005) estableció para la identificación de la coordinación: la coordinación debe volver a la primera frase en la segunda, su entonación debe ser modal de frase independiente para ambas y debe formar una serie abierta (en la que la segunda frase sigue a la primera, la tercera a la segunda... y así sucesivamente). El objetivo fundamental de este tipo de enunciados es el de “poner en (oscura) relación lo que se dice con lo que se viene diciendo, facilitando así el fluir discursivo de los interlocutores” (Vigara, 2005, p. 125).

Un ejemplo sería: “[...] a mí me puede explicar lo que quiera, que me lo creo, ... Claro, ¡si yo no tengo ni idea de las leyes de Kepler! Me suenan ¿no? ... *Pero* este hombre se puso allí a explicar, hasta que se levantó un tío y dijo: «pero ¿oiga que usted qué está haciendo ahí?»” (Vigara, 2005, p. 125). En este caso, no tomaremos la *y* como nexo inespecificativo por las dudas que surgen con respecto a la suspensión de su funcionamiento gramatical como conjunción copulativa.

2. La subordinación inespecificativa: QUE, PORQUE, PUES.

Asimismo, también se pueden encontrar nexos subordinantes que se usan con un nuevo sentido, diferente al que se les suele asignar. Al perder su función gramatical, adquieren un significado conversacional y discursivo. En algunos casos, la subordinación inespecificativa se emplea para retomar el hilo comunicativo y seguir avanzando en la conversación. Según Vigara, el nexo en este tipo de enunciados es enfático y no funciona como nexo de proposición subordinada. Por ejemplo, el nexo *porque* no siempre expresa la causa de lo dicho en el enunciado, sino que también puede manifestar la razón subjetiva del hablante. Un ejemplo de esto sería: “Claro, pues entonces ¿por qué nos engañan? ... claro, *porque* nos están engañando”. (Vigara, 2005, p. 127). Se trata de una subordinación claramente inespecificativa por la ausencia del verbo principal explícito y el valor discursivo de la unidad en cuestión.

3. La presencia de los sujetos de la enunciación en el discurso: TÚ, YO.

Autores como M. Criado, M. Seco y A. M. Vigar, han caracterizado frecuentemente el español coloquial como egocéntrico (Briz, 2001). La frecuente actualización en la conversación coloquial hace que el YO-AQUÍ-AHORA (relacionado con el TÚ presente) sea el centro deíctico personal, espacial y temporal (Briz, 1998); ésta es la orientación subjetiva de la que parten las expresiones deícticas.

Debido a este carácter egocéntrico de la conversación coloquial, existe una tendencia a construir la oración a partir de (y entorno a) un elemento personal que se suele destacar al principio (Vigar, 2005). La autora afirma que la presencia del *yo* o el *tú* en el discurso de registro coloquial lo hace más expresivo y su objetivo se convierte en persuadir al interlocutor. Es por ello por lo que el *sujeto de la enunciación* (que no debe confundirse con el *sujeto del enunciado*) es uno de los elementos que se focaliza con más frecuencia. Además, si en la lengua coloquial (con una dirección de un “yo” hablante a un “tú” interlocutor) estos sujetos aparecen explícitos, siempre serán enfáticos, redundantes y, como hemos dicho antes, expresivos. Antonio Briz (2001), consideró que el empleo del *yo* como voz principal del diálogo tiene un fin estratégico de intensificación o atenuación por parte del hablante. Éste maximiza o minimiza su presencia en el discurso y su papel en la conversación para alcanzar sus objetivos. Si decide intensificar el enunciado, el emisor maximizará su papel en el mismo, empleando el *yo* para realzar su presencia y elevar (y proteger) su imagen frente al otro. En cambio, si su objetivo es atenuar su mensaje, el individuo optará por sustituir el *yo* por un verbo opinativo o de decir; por ejemplo: “Yo creo / A mí me parece”¹⁸ (Briz, 2010, p. 56). Por otro lado, el *tú* aparece casi siempre de forma directa y solo algunas veces de forma atenuada (cuando hay una estrategia detrás) (Briz, 2001).

¹⁸ Cabría indicar que, si bien es necesario reconocer el valor atenuador del verbo realizativo empleado, encontramos que el sujeto de la enunciación sigue presente en el enunciado, independientemente de que se emplee el pronombre personal o no. De esta forma, el valor atenuador del verbo aparece equilibrado con el poder intensificador de la presencia del sujeto de la enunciación (presente, por ejemplo, en la propia desinencia del verbo empleado).

5.2.2. Marcadores pragmático-discursivos:

Los marcadores pragmático-discursivos son los siguientes:

1. Formular. Reformular y avanzar: ENTONCES, BUENO, VENGA, ASÍ QUE.

Como ya hemos visto en el capítulo anterior, cuando conversamos con alguien, también estamos llevando a cabo una actividad argumentativa, que va siempre unida a la actividad formulativa (Briz, 2010). “Formular significa ir resolviendo los muchos problemas comunicativos que se plantean a lo largo de una interacción cara a cara, más si ésta es una conversación coloquial” (Antos, 1982 y Chafe, 1982, citados en Briz, 2010). Estos problemas comunicativos se dan porque los interlocutores no saben previamente cómo va a desarrollarse su discurso (debido a la inmediatez y a la actualización propias de la conversación).

Para resolver estos posibles problemas de organización o comprensión del mensaje, el hablante hace uso de estos marcadores. Según el autor, el esfuerzo que emisor y destinatario deben ejercer a la hora de codificar y descodificar los mensajes se ve reflejado en sus enunciados. Es por ello por lo que sirven para estructurar y organizar el discurso.

Cuando un hablante usa *bueno* manifiesta su intención de volver sobre lo que ha dicho anteriormente para matizarlo. En ocasiones, *bueno* puede funcionar también como partícula de puntuación, a modo de pausa oralizada.

El hablante, al recurrir a estos marcadores, manifiesta su intención de evitar contradicciones en su discurso y a su vez que éste presente un hilo argumentativo lógico y de fácil interpretación para su interlocutor.

Además, según Briz (2010), para hacer que el discurso avance, el individuo puede usar marcadores como *entonces* o *venga*. El uso de *entonces* como continuador de actos sirve para dotar el discurso de orden y progresión y puede ser utilizado por el hablante para recuperar un tema anterior que ha sido interrumpido por alguna digresión o por una pérdida de turno. Cuando *así que* funciona como recapitulador de lo dicho, se puede sustituir por *entonces*.

2. Control del mensaje (inicio, progresión, cierre): PUES, VALE, EN FIN, O SEA, BIEN, POR CIERTO.

Este tipo de marcadores se utilizan para señalar las partes del discurso y, por tanto, para ordenar la materia discursiva. Asimismo, el autor afirma que regulan el inicio, la progresión y

el cierre de los turnos y mensajes de los interlocutores. El conector *pues* acota y realza las partes del discurso con el fin de mantener el hilo del mensaje. Además, puede funcionar como marcador de reanudación y de reformulación. En otras ocasiones, *pues* refuerza un acto ilocutivo de afirmación o de negación (“*Pues sí, pues no*”) e incluso sirve como iniciador de una réplica por la que el hablante muestra su acuerdo o desacuerdo con lo dicho (“*Pues, ¿quién es para decirme que me calle?*”) (Briz, 2010, p. 210).

Los marcadores que regulan el inicio del mensaje suelen aparecer al principio de una intervención reactiva o de una secuencia de la conversación. Consisten en pausas que permiten al hablante ganar un poco de tiempo para pensar y “planear” lo que se va a decir a lo largo de la conversación. Es decir, son unidades que rellenan los huecos que se producen cuando el hablante comienza su discurso o no encuentra la manera de proseguir o no sabe qué responder (Valdés, 1944; Covarrubias. 1977, citados en Briz, 2010). Un ejemplo de marcador de inicio sería el conferenciante que usa *bien* al comienzo de su intervención (Briz, 2010).

Las marcas de progresión otorgan continuidad al discurso. Marcadores como: *o sea* o, *por cierto* funcionan como *reformuladores* del discurso, ya que el hablante los emplea para rectificar, precisar o explicar un tema, una actitud o incluso un acto argumentativo. Los reformuladores del tipo *o sea, es decir...* marcan la relación de equivalencia comunicativa entre las dos partes del enunciado: el de partida (el que es explicado) y el de llegada (el que explica). Por otro lado, *por cierto* tiene una función digresiva que permite al hablante introducir una nueva idea o cambiar de tema. Estas digresiones, según Briz (2010), serán consentidas por el interlocutor si están vinculadas a una idea anterior o la situación y circunstancias de la conversación. Si no están motivadas o el interlocutor no logra encontrar el vínculo, éste pasará a demandar más información. Conversacionalmente, la función digresiva de *por cierto* puede también manifestar un tono recriminatorio.

Las marcas de cierre (*vale, en fin*) pueden ser utilizadas para cerrar o dar por terminada la conversación.

Todos estos marcadores son unidades polivalentes por lo que, en ocasiones, funcionan en diferentes niveles de la organización jerárquica del mensaje. Por tanto, para conocer su valor particular se deben presentar en el contexto en el que se usan (p. 221). El valor de estas unidades dependerá de: su naturaleza léxica, su emplazamiento sintagmático y su entorno prosódico.

3. Control del contacto: MIRA, ESCUCHA, FÍJATE, OYE(S), SABES, ENTIENDES, VES, COMPRENDES, HOMBRE, LA VERDAD.

Los marcadores que sirven para controlar el contacto tienen una función interpersonal y socializadora. Esta función, según Briz (2010), manifiesta constantemente el tipo de relación que existe entre emisor y destinatario y sus enunciados. Marcadores como: *mira, escucha, fíjate, oye(s), sabes, entiendes, ves, comprendes, hombre, la verdad...* contribuyen a la transmisión de un objetivo ilocutivo asociado con las funciones expresivo-apelativa y fática. Además, aparecen en el discurso por diversas razones: para reforzar o justificar los razonamientos y argumentos del hablante ante su interlocutor, para llamar la atención de este y así mantener su contacto o simplemente como fórmulas exhortativas y apelativas para conseguir la implicación del destinatario. Cuando el hablante utiliza estos marcadores puede manifestar su desacuerdo con su interlocutor e incluso llegar a exigirle un cambio de actuación. En definitiva, son unidades modales que funcionan como refuerzos del acto ilocutivo implícito (protesta, orden, recriminación, advertencia...). Al igual que los marcadores de control del mensaje, el valor comunicativo de los marcadores de control del contacto depende de su valor léxico, su posición y, su entonación.

4. Expresiones de subjetividad: ¡!, SIGNOS DE PUNTUACIÓN, EMOTICONOS, REPETICIONES VOCÁLICAS.

Las expresiones de subjetividad son marcas a las que debemos prestar atención pues se emplean con profusión en el canal de comunicación seleccionado para nuestro estudio.

Las repeticiones vocálicas o alargamientos fónicos aparecen en el discurso por varias razones. Primeramente, el hablante puede emplear estas marcas debido a su escasa destreza lingüística y a la necesidad que tiene de tener un apoyo para ganar tiempo y pensar lo que va a comunicar (Briz, 2010). Por otro lado, las repeticiones vocálicas también pueden actuar como refuerzos de lo dicho e intensificar lo que el emisor quiere transmitir. Por ejemplo, en esta conversación, el emisor 1 intensifica sus ganas de fumar mientras que el emisor 2 está recriminando lo dicho anteriormente.

E1: ¡Uff! *Fumabaa*.

E2: ¡*Dejaloo!*¹⁹

¹⁹ Briz, 2010.

Pero también pueden servir como atenuadores en casos como:

A: ¿qué tal los nuevos planes de estudio?

B: *bueeno* no están mal.

Los emoticonos, según Agnese Sampietro (2016) son “representaciones esquemáticas de expresiones faciales o pequeñas imágenes que se añaden a los mensajes electrónicos” (p. 15). Su estudio científico se ha ido desarrollando a la vez que lo han hecho también las tecnologías digitales. Además, la autora defiende que este auge actual de los emoticonos refleja las características de la cultura digital actual²⁰. Es por ello por lo que Sampietro (2010) afirma que, al introducir un emoticono en el mensaje, el hablante puede informarnos sobre su relación interpersonal con su interlocutor (entre otras cosas). Los emoticonos, también pueden trazar perfiles de los hablantes que los usan, por este motivo los incluimos en nuestra investigación. Según los estudios, las mujeres parecen utilizar más emojis en wasap que los hombres²¹. Este hecho, ha intentado ser explicado por algunos autores mediante afirmaciones como: las mujeres son más propensas a expresar emociones²² o ellas tienen un sentido estético superior a los hombres²³. Sin embargo, creemos que estas razones caen de lleno en los estereotipos sociales. Por otro lado, también es común asociar un uso mayor de los emoticonos a grupos de edad joven, más que a los hablantes de edades avanzadas²⁴. Esta última tendencia podría justificarse por el uso habitual y generalizado de las nuevas tecnologías por parte de los jóvenes y por ser ellos el motor del cambio lingüístico²⁵. Otro dato destacable que Sampietro (2016) comenta en su tesis es que dependiendo de la lengua en la que el usuario está escribiendo su mensaje, usará un número mayor o menor de emoticonos y emojis. Cuando escribe en una segunda lengua (L2), el emisor utilizará una cantidad superior de emoticonos (que cuando escribe en su lengua materna) para intentar paliar la distancia emocional que se tiene al hablar en la lengua extranjera²⁶ o incluso para suplir sus carencias verbales en ella²⁷. Asimismo, la aparición de los emoticonos en los intercambios comunicativos suele ser más frecuente en contextos dialógicos²⁸ y por tanto no suelen abrir la

²⁰ Adami y Kress, 2010, citados en Sampietro, 2016.

²¹ Pérez Sabater, 2015, citada en Sampietro, 2016.

²² Komrsková, 2015, Nishimura, 2015, Parkins, 2012, citados en Sampietro, 2016.

²³ Witmer y Katzman, 1997, citados en Sampietro 2016.

²⁴ Azuma y Maurer, 2007, Baron y Ling, 2001, Betti, 2003, Thurlow y Brown, 2003, citados en Sampietro, 2016.

²⁵ Stockton, 2015, citado en Sampietro, 2016.

²⁶ Aragon, Chen, Kroll y Feldman, 2014, citados en Sampietro, 2016.

²⁷ Halvorsen, 2012, Vandergriff, 2014, citados en Sampietro, 2016.

²⁸ Kavanagh, 2010, Halvorsen, 2012, citados en Sampietro, 2016.

conversación, sino que se colocan en el texto a modo de pausa o signo de puntuación o al final de la oración²⁹. El empleo de emoticonos tiene, por tanto, un papel fundamental en la pragmática ya que algunos autores sostienen la idea de que estas pequeñas representaciones faciales o imágenes pueden llegar a asumir las tareas que desempeñan los signos de puntuación tradicionales o actuar como marcadores de cambio de turno.³⁰

Esto llevó a los investigadores a pensar que el uso de los emoticonos es preferentemente pragmático (Sampietro, 2016). Este carácter pragmático de los emoticonos se traduce en tres funciones diferentes³¹:

1. Función indicadora de afectos y emociones vividas por el emisor.
2. Función irónica o humorística.
3. Función estratégica de cortesía con el fin de atenuar actos potencialmente amenazadores de la imagen del destinatario.

Los emoticonos además de ser marcadores emocionales y expresivos también sirven para interpretar los enunciados correctamente (y desambiguarlos), indicar la relación que el emisor quiere mantener con el destinatario y moderar el carácter negativo del mensaje³². Calero Vaquera (2014, citado en Sampietro, 2016) también considera que los emoticonos tienen tres funciones fundamentales:

1. Ayudan a expresar ideas y sentimientos.
2. Enfatizan el mensaje.
3. Aclaran contenidos implícitos como la ironía.

Por tanto, autores como Dresner y Herring (2020, citados en Sampietro, 2016) concluyeron en su estudio que los emoticonos son indicadores de la fuerza ilocutiva del enunciado al que acompañan facilitando así la correcta interpretación del destinatario.

En contextos de socialización, los emoticonos suelen estar más presentes que en contextos relacionados con la realización de tareas o el cumplimiento de objetivos³³. De esta manera, se puede llegar a concluir que el uso de emoticonos en contextos profesionales podría ser inadecuado, aunque, autores como Gettinger y Koeszegi (2015, citados en Sampietro, 2016) defienden que es más probable que dos personas lleguen a un acuerdo si emplean emoticonos en sus mensajes. El estudio realizado por Walther y D'Addario en 2001³⁴ demostró que los

²⁹ Provine et al., 2017, citados en Sampietro, 2016.

³⁰ Markman y Oshima, 2007, Vela, Delfa y Jiménez Gómez, 2011, citados en Sampietro, 2016.

³¹ A. Wilson, 1993, citado en Sampietro, 2016.

³² Marcoccia, 2000, López Quero, 2013, Crystal, 2002, Galán, 2002, citados en Sampietro, 2016.

³³ Derks, Bos y Gumbkow, 2007, citados en Sampietro, 2016.

³⁴ Citados en Sampietro, 2016.

mensajes en los que se usaban emoticonos (coherentes con su contenido) poseían un valor emotivo superior que los que no los contenían, por lo que se puede deducir que los emoticonos son potenciadores del valor emotivo de los mensajes. Finalmente, podemos afirmar que el uso pragmático de los emoticonos facilita la gestión de la interacción a los interlocutores (Sampietro, 2016).

5. Expresiones atenuadoras: CREO, PIENSO, ME PARECE

La atenuación “[...]” consiste en no expresar todo lo que se quiere dar a entender, sin que por esto deje de ser bien comprendida la intención del que habla [...]”³⁵. Cuando un hablante quiere atenuar la fuerza ilocutiva de un acto (asertivo o exhortativo) puede recurrir a la llamada *atenuación pragmática performativa* que suaviza el acto de habla correspondiente. Según el autor, para llevar a cabo este tipo de atenuación, el emisor elegirá un verbo performativo. Los verbos performativos se caracterizan por estar en presente de la primera persona del singular en voz activa. Enunciados como “*Creo, pienso, me parece que...*” serían claros ejemplos de este tipo de expresiones atenuadoras. Antonio Briz (2010) pone de ejemplo la siguiente frase: “Yo pienso que tampoco me he portado tan mal” (p. 151)³⁶.

6. Fórmulas rutinarias: Y PUNTO, Y NADA, Y YA ESTÁ, ¿QUÉ TE IBA A DECIR?, Y ESO, Y TAL, ¡QUÉ VA!, DESDE LUEGO, PUES ESO, HASTA LUEGO, ¿QUÉ HAY?, ¿QUÉ TAL?, ¿QUÉ PASA?, ¿CÓMO TE VA LA VIDA?, ¿QUÉ TE CUENTAS?, ¡QUÉ BONITO!, ¿QUÉ ME DICES?, ¡QUÉ FUERTE!, ¡QUÉ PASADA!, ¡QUÉ PUTADA!, ¡HIJA MÍA!, PUES VAYA, ¿QUÉ QUIERES QUE TE DIGA?

Las fórmulas rutinarias son un tipo de unidad fraseológica. Según María Belén Alvarado Ortega (2010) una unidad fraseológica es “la combinación formal y psico-lingüísticamente estable de unidades léxicas compuestas por dos o más palabras que forman parte de la competencia léxica de los hablantes, cuyo límite superior se encuentra en el nivel de la oración compuesta y que, a su vez, puede poseer idiomatidad” (p. 19). La autora afirma que las fórmulas rutinarias se encuentran dentro de los enunciados fraseológicos. Estos

³⁵ RAE, 1992, 21ª edición, citado en Briz, 2010.

³⁶ Este tipo de actos de habla en los que el emisor trata de atenuar su impacto en su destinatario son los que Brown y Levinson denominan *face threatening acts* (FTA's) o los actos amenazadores de la imagen. Para saber más véase a N. Campos (2018).

enunciados “[...] funcionan como secuencias autónomas de habla, su enunciación se lleva a cabo en unidades de entonación distintas; en otras palabras, son unidades de comunicación mínimas” (Zuluaga, 1980, citado en Alvarado, 2010, p. 22). Las fórmulas rutinarias son unidades sin autonomía textual (dependen del contexto en el que se producen) que poseen un significado social, expresivo o discursivo y se pueden distinguir entre³⁷:

- 1) Las fórmulas discursivas, que organizan el discurso. Entre ellas destacan las de apertura y cierre de la conversación (*¿Qué hay?*³⁸) y las de transición, que regulan el curso de la interacción (*a eso iba*).
- 2) Las fórmulas psico-sociales, que funcionan para expresar el estado mental y los sentimientos. Destacan las expresivas (*lo siento*), las comisivas (*palabras de honor*), las directivas (*¡al grano!*), las asertivas (*te lo digo yo*), las rituales (*buenos días*) y las misceláneas (*pelillos a la mar*).

Alvarado (2010) afirma que las fórmulas rutinarias al ser unidades fraseológicas compuestas por dos o más palabras ritualizadas, su límite superior se encuentra, como ya hemos dicho anteriormente, en la oración compuesta. Además, añade que la fijación, la idiomatización y la independencia son algunos de los rasgos que caracterizan a las fórmulas rutinarias. La autora también señala que son estructuras estables tanto en su forma como en su significado, pues se encuentran en el léxico de todos los hablantes de una comunidad lingüística. Es por ello por lo que los hablantes las emplean para expresar su opinión sobre algo o como apoyo lingüístico que facilite el transcurso de la conversación (Alvarado, 2010).

Teniendo claro el concepto de fórmula rutinaria, pasaremos ahora a comentar el significado de algunas de ellas mediante ejemplos en los que podremos ver su contexto de uso.

- Y TAL, Y ESO, Y NADA:

H: ¿Te acuerdas de Alaska y Dinarama?

A: Me acuerdo del nombre y *tal*, pero no de la música.

En este caso, y *tal* está sustituyendo información que el emisor considera que no es relevante porque ya es conocida por su destinatario. Aunque esta fórmula rutinaria también puede ser empleada en otros contextos como conclusivo de enumeración (“Me gustan más los colores cálidos, el rojo, el naranja, el amarillo... y *tal*”).

A pesar de que la fórmula rutinaria y *eso* también pueda presentar el mismo valor de y *tal* (supresión de conocimientos compartidos) no se tratan de variantes de una misma fórmula

³⁷ Corpas, 1996, citada en Alvarado, 2010.

³⁸ Todos los ejemplos han sido sacados del libro de Alvarado Ortega (2010).

pues pueden llegar a presentar valores distintos en otros contextos. *Y eso* puede funcionar en contextos en los que el hablante desea concluir lo que está enunciando. Este mismo valor lo adquiere la fórmula rutinaria y *nada* en el siguiente ejemplo:

H2: ¡Ah! <ininteligible>. Pues, nada, muy mal. Lo pasé muy mal... y es cuando más falta te hacen ahí las amigas o los familiares o tal... y las noches es lo peor, o sea, es... en la vida lo he pasado tan mal... Las noches son interminables, el día larguísimo... y *nada*, eso³⁹.

Es por ello por lo que no funcionarían como variantes ya que estas deben presentar la misma función en todos los contextos (Alvarado, 2010).

- ¡QUÉ VA!:

A: Pero ¿Roberto entiende de relojes?

B: ¡Qué va! No tiene ni idea.

Como podemos ver en este ejemplo, *¡qué va!* funciona como elemento para indicar negación ante la pregunta formulada.

- ¡QUÉ BONITO!:

En el siguiente ejemplo, varios amigos discuten sobre la legalización de las drogas.

H3: Pues lo veo perfecto. Además, las drogas son <vacilación> muy buenas para combatir el dolor... la ansiedad... el sufrimiento... <simultáneo> la desidia.

E1: *¡Qué bonito!* </simultáneo> Además lo que dijeron también el otro día en las noticias... lo que dijeron también el otro día en las noticias... que <fático = duda> <estilo indirecto> el chocolate lleva una sustancia, que la lleva también la maría... la marihuana... una sustancia... que, por eso, la gente cuando esta deprimida come chocolate y se sienten <sic> mejor </estilo indirecto>⁴⁰.

E1 responde a la intervención de H3 con *¡qué bonito!* para mostrar su desacuerdo con respecto a su afirmación, por lo que esta fórmula rutinaria implica un valor negativo ya que es irónica.

Por otro lado, esta fórmula rutinaria puede llegar a adquirir un valor positivo (además de mostrar complicidad) en otros contextos como en esta conversación en la que se habla sobre la situación laboral de nuestro país:

H1: Depende mucho la valoración que la gente quiera dar, o sea, está claro que el enchufe sigue existiendo y ¡que l tiene más fácil los que tienen pelas, pa<(r)><(a)> que nos vamos a

³⁹ Ejemplo tomado por Alvarado (2010) del corpus COVJA (1997), grupo 2, p. 85.

⁴⁰ Ejemplo tomado por Alvarado (2010) del corpus COVJA (1997), grupo 10, p. 321.

engañar!, es verdad, pero si tú quieres y, aunque no tengas un duro, estés viviendo bajo un puente, si tú te lo curras llegarás hasta donde te propongas.

H3: ¡*Qué bonito!*, ¡ojalá!

H1: Es que es así⁴¹.

- PUES VAYA:

En el caso de esta conversación telefónica, dos amigos hablan de como uno ha llegado al piso del otro. La fórmula *pues vaya* expresa decepción ante lo que G le está contando.

G: si/ yo sabía que eraa/ este piso yy bueno pues// por lo que me acuerdo yo de orientación y tal sabia que mas o menos era// aquí ¿no?// y he llamao y como no abría nadie yo digo a lo mejor no es aquí// y he llamao ahí al la(d)o y tampoco estaban.

E: ¿sí?

G: *pues vaya*⁴².

- DESDE LUEGO:

Esta fórmula rutinaria de tipo evidencial expresa una modalidad epistémica. Esto significa que sirve para marcar el grado de certeza o conocimiento del hablante que la usa con respecto a lo dicho anteriormente (Alvarado, 2010).

- ¿QUÉ QUIERES QUE TE DIGA?, ¡QUÉ PUTADA!, ¡HIJA MÍA!, ¡QUÉ FUERTE!,
¡QUÉ PASADA!, ¿QUÉ ME DICES?:

Estas fórmulas rutinarias son de tipo subjetivo porque con ellas, los hablantes expresan juicios de valor. En concreto, se tratan de formulas rutinarias subjetivas afectivas. Todas manifiestan una reacción emocional: decepción (*¿Qué quieres que te diga?*), solidaridad (*¡Qué putada!*, *¡Hija mía!*) y sorpresa (*¡Qué fuerte!*, *¡Qué pasada!*, *¿Qué me dices?*). Con este tipo de fórmulas, el emisor trata de expresar su actitud y sus emociones a su destinatario (Alvarado, 2010). Un ejemplo sería:

C: [y yo te dije] dale en to 'l cogote/ y él me oyó se giró y en vez del cogote en [todo el ojo].

A: [en todo el ojo que le dio].

B: [(RISAS) *¡qué putada!* (RISAS)]⁴³.

⁴¹ Ejemplo tomado por Alvarado (2010) del corpus COVJA (1997), grupo 10, p. 296.

⁴² Ejemplo tomado por Alvarado (2010) del *corpus de conversaciones coloquiales* de Briz y el grupo Val.Es.Co. (2002), p. 83.

⁴³ Ejemplo tomado por Alvarado (2010) del *corpus de conversaciones coloquiales* de Briz y el grupo Val.Es.Co. (2002), p. 64.

En este caso con *¡qué putada!* el emisor B, aunque transmite solidaridad, pena y complicidad con los interlocutores no le es posible ocultarles sus risas.

- ¿QUÉ TE CUENTAS?, ¿QUÉ TAL?, ¿QUÉ HAY?, ¿QUÉ PASA? ¿CÓMO TE VA LA VIDA?, HASTA LUEGO, Y YA ESTÁ, Y PUNTO, ¿QUÉ TE IBA A DECIR?, Y ESO, Y TAL, PUES ESO

Todas estas fórmulas rutinarias se engloban dentro del grupo de las discursivas. Se emplean para organizar, estructurar y hacer que la conversación avance (Alvarado, 2010). Las fórmulas rutinarias discursivas se pueden dividir en dos grupos:

1) Fórmulas rutinarias discursivas de apertura y cierre de la conversación:

Este tipo de fórmulas organizan el discurso y codifican las relaciones sociales de los hablantes en la conversación (Corpas, 1996, citada en Alvarado, 2010). Muchas de ellas, se relacionan con la cortesía pues el hablante las emplea estratégicamente para evitar las posibles amenazas a la imagen⁴⁴ de su interlocutor. Las fórmulas de apertura que hemos seleccionado para nuestro estudio son: *¿qué te cuentas?*, *¿qué tal?*, *¿qué hay?*, *¿qué pasa?*, *¿cómo te va la vida?* Y la de cierre es: *hasta luego*.

En este ejemplo, varias personas hablan sobre la ubicación del piso de otra (ausente en la conversación) cuando de repente aparecen otros dos individuos:

A: = donde está ContiNENTE // donde está Continente// huerta y huerta y huerta.

Aparecen el hijo (P) y la esposa de A.

G: ¡holaaa!// [°(saluda a tu padre)°].

G: [*¿qué hay?* *¿cómo estás?*].

P: bien.

J: [*¿ya te vas de fiesta?*]⁴⁵.

El emisor G muestra cortesía ante las personas que acaban de llegar para integrarlos en la conversación (Alvarado, 2010).

2) Fórmulas rutinarias discursivas de transición:

Las fórmulas rutinarias de transición también estructuran el discurso además de controlar la fluidez de los intercambios (Corpas, 1996, citada en Alvarado, 2010). Su función es discursiva, es decir, toman su valor en el contexto en el que se producen y su significado viene dado por el papel que tengan en la estructuración de los intercambios de la conversación

⁴⁴ Los "Face threatening acts" de Brown y Levinson que ya comentamos anteriormente.

⁴⁵ Ejemplo tomado por Alvarado (2010) del *corpus de conversaciones coloquiales* de Briz y el grupo Val.Es.Co. (2002), p. 171.

(Alvarado, 2010). Dentro de este tipo, hemos seleccionado fórmulas que sirven para concluir: *y ya está*, *y punto*, otras que sirven para conservar o solicitar el turno de palabra: *¿qué te iba a decir?* y finalmente las que sirven para omitir información: *y eso*, *y tal*, *pues eso*.

En el siguiente ejemplo, L y E charlan sobre los ligues de una noche:

L: no pero dentro de lo que cabe sí porque eso.

E: pero el hecho de que tú rechaces esa situación denota seguridad ¿o no? Tú estás totalmente segura de ti y de lo que quieres de que tienes una relación estable *y punto*⁴⁶.

6. Desarrollo práctico:

6.1. Hipótesis de trabajo y metodología.

6.1.1. Tests de Hipótesis y prueba Chí cuadrado.

6.2. Análisis del corpus.

6.2.1. Análisis Chí Cuadrado.

6.2.2. Tabla de error.

6.2.3. Análisis perfiles lingüísticos.

En este capítulo explicaremos en qué consisten los tests de hipótesis, y en concreto hablaremos sobre la prueba chí cuadrado ya que es la que hemos usado para nuestro estudio. Finalmente, presentaremos los resultados obtenidos en el programa UAM Corpus Tool, el cual nos ha sido de gran ayuda para analizar nuestro corpus y obtener las conclusiones finales.

6.1. Hipótesis de trabajo y metodología:

Nuestra hipótesis de trabajo es que las marcas identificadoras del registro coloquial en español son útiles para la identificación de autores. Es decir, las marcas seleccionadas que se dan en el entorno wasap van a tener un comportamiento similar en diferentes textos producidos por el mismo autor. Por el contrario, cuando comparemos autores diferentes vamos a observar que existen mayores diferencias en el uso de esas marcas.

A continuación, pasaremos a describir las muestras que hemos tomado para la investigación. Nuestro corpus se constituye de un conjunto de textos agrupados por autoría, en concreto 50. Disponemos de 10 textos de autoría conocida y un conjunto de textos agrupados que hemos

⁴⁶ Ejemplo tomado por Alvarado (2010) del *corpus de conversaciones coloquiales* de Briz y el grupo Val.Es.Co. (2002), p. 98.

denominado generales para poder comparar el comportamiento lingüístico de cada autor conocido con los comportamientos del resto de hablantes y así poder demostrar si es posible identificar a un autor con las marcas pragmáticas que hemos seleccionado en el estudio. Estos textos generales pertenecen a un grupo de 40 informantes. Los 10 textos restantes han sido producidos por un total de 5 autores conocidos. Todos los textos del corpus han sido producidos en la aplicación de mensajería instantánea wasap. Bernal, Riveiros y Garayzábal (2014) consideraron que los textos que se producen a través del correo electrónico, las redes sociales y las aplicaciones de mensajería del móvil a parte de suponer un reto para los lingüistas forenses (por sus características especiales como los emoticonos, la concisión de contenidos o su corta extensión), son también ser de vital importancia como pruebas para la resolución de casos criminales como el de Danielle Jones⁴⁷. Las muestras empleadas han sido extraídas del corpus ForenUca. En este corpus, solo contiene una parte de la conversación pues se necesitaba la autorización de todos los participantes de la interacción. Es por ello por lo que tenemos los textos de wasap sin el interlocutor. El contenido de las conversaciones trata sobre asuntos de la vida cotidiana y diaria. Finalmente, el número de palabras de cada texto es el siguiente: 1.714, 3.062, 1.935, 850, 1.903, 530, 1.294, 1.662, 1.285, 1.301, 2.012, 1.498, 2.039, 2.182, 2.535, 1.486, 2.304, 1.360, 2.045, 693, 2.717, 1.756, 1.843, 1.235, 956, 2.826, 2.097, 1.197, 1.380, 579, 2.555, 620, 1.968, 1.972, 2.045, 1.127, 2.244, 2.046, 1.288, 1.805, 542, 317, 1.567, 2.261, 524, 1.626, 3.015, 767, 2.046, 1.731.

En cuanto a la metodología utilizada para el desarrollo práctico, el programa UAM Corpus Tool, creado por el lingüista computacional Mick O'Donnell, nos ha sido de gran utilidad como herramienta de observación. Este programa, comprende un conjunto de herramientas que permiten al lingüista anotar lingüísticamente un texto de manera manual o semiautomática. Además, es posible realizar búsquedas de palabras o características dentro de un corpus proporcionando un análisis estadístico de los datos (http://www.linguisticsweb.org/lib/exe/fetch.php?media=linguisticsweb:tutorials:linguistics_tutorials:manual_annotation:uam_corpus_tool.pdf) Este análisis estadístico viene expresado por dos tipos de pruebas estadísticas entre ellas la prueba de chí-cuadrado.

⁴⁷ Un mensaje de texto resolvió el caso de Danielle Jones. Una chica de 15 años que desapareció en junio de 2001 y fue asesinada. Su tío Stuart Campbell fue su secuestrador y asesino. El acusado, presentó en su defensa una serie de mensajes de texto que supuestamente Danielle había enviado tras desaparecer. Gracias al análisis del Profesor Coulthard salieron a la luz las diferencias en la escritura de palabras antes y después de la desaparición de Danielle. Coulthard se dio cuenta de que Danielle escribía la palabra “what” de manera diferente a cuando lo hacía antes de ser secuestrada. En uno de los supuestos mensajes que Stuart presentó, la palabra estaba escrita como “wot” cuando el análisis de Coulthard determinó que la chica solía escribirla como “wat”. Finalmente, Campbell fue declarado culpable de su asesinato en 2002 (Bernal, Riveiros, Garayzábal, 2014).

6.1.1. Tests de Hipótesis y prueba Chí cuadrado:

Para saber cómo funciona la prueba chí-cuadrado debemos saber antes en qué consisten los tests de hipótesis, ya que chí-cuadrado es uno de ellos.

Los tests de hipótesis son un procedimiento estadístico que consiste en investigar si una hipótesis planteada por un investigador sobre una característica de una población es falsa o verdadera basándose en la evidencia que proporciona una muestra de datos. Para ello, la prueba de hipótesis examina dos afirmaciones opuestas sobre la población que se está estudiando conocidas como: la hipótesis nula y la hipótesis alternativa. La hipótesis nula es la que se está comprobando en el estudio y, por tanto, es un enunciado que afirma que “no hay diferencias/efectos” (<https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/basics/what-is-a-hypothesis-test/>).

Por el contrario, la hipótesis alternativa es el enunciado que el investigador quiere demostrar y concluir y por tanto afirma que “sí hay diferencias/efectos” basándose en la evidencia que dan los datos de la muestra (<https://www.addlink.es/noticias/minitab/2852-que-es-una-prueba-de-hipotesis>). De esta manera, los tests de hipótesis determinan cuándo rechazar la hipótesis nula o H_0 y cuándo la hipótesis alternativa o H_1 es verdadera o por el contrario cuándo aceptar la hipótesis nula y descartar la alternativa.

Según Triola (2009) chí cuadrado “se basa en las diferencias entre valores observados y esperados, de manera que una concordancia cercana entre los valores observados y esperados conducirá a un valor de chí cuadrado pequeño y un valor de P grande. Una discrepancia grande entre los valores observados y esperados conducirá a un valor de chí cuadrado grande y un valor P pequeño” (p. 594). Cuando el valor de chí cuadrado es pequeño y el valor de P es grande, la hipótesis nula no se rechaza. Sin embargo, cuando chí cuadrado tiene un valor grande y P tiene un valor pequeño, se rechaza la hipótesis nula.

La fórmula de la prueba chí-cuadrado es la siguiente:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

El análisis se basa en comparar los resultados que el programa UAM Corpus Tool nos muestra estadísticamente con la prueba chí cuadrado. Cuando el programa detecta una diferencia significativa estadísticamente entre dos textos, el programa lo detecta y muestra un

símbolo que se compone de dos pequeñas cruces⁴⁸ amarillas (cuando la probabilidad estadística de que provengan de un mismo autor es menor que 0.05) o incluso tres cruces⁴⁹ (cuando es menor que 0.01).

Partimos entonces de la base de que un mismo autor tenderá a repetir sus marcas en sus escritos de una manera proporcional. El test chí cuadrado nos muestra si las diferencias observadas son estadísticamente significativas como hemos advertido en el párrafo anterior. Para completar esta tarea de una forma más eficiente, las comparaciones se han hecho de manera proporcional debido a que los textos no tenían el mismo tamaño. La herramienta UAM Corpus Tool permite simular que los todos textos tuvieran unas 1000 palabras. El ajustar su proporción hace que los resultados sean más fiables.

6.2. Análisis del corpus:

En primer lugar, debemos anotar todos los textos con marcas coloquiales para poder observar así cómo se comportan en los diferentes escritores. Nuestro análisis se basa en observar en los textos las siguientes marcas:

1) Sintácticas

Coordinación inespecificativa: Y, PERO

Subordinación inespecificativa: QUE, PORQUE, PUES.

Presencia de los sujetos de la enunciación en el discurso: TÚ, YO.

2) Conectores pragmático-discursivos:

Formular. Reformular y avanzar: ENTONCES, BUENO, VENGA, ASÍ QUE.

Control del mensaje (inicio, progresión, cierre): PUES, VALE, EN FIN, LA VERDAD, HOMBRE, O SEA, BIEN, POR CIERTO.

⁴⁸ Medium significance.

⁴⁹ High significance.

Control del contacto: MIRA, ESCUCHA, FÍJATE, OYE(S), SABES, ENTIENDES, VES, COMPRENDES.

3) Expresiones de subjetividad: ¡!, SIGNOS DE PUNTUACIÓN, EMOTICONOS, REPETICIONES VOCÁLICAS.

4) Expresiones atenuadoras: CREO, PIENSO, ME PARECE.

5) Fórmulas rutinarias: Y PUNTO, Y NADA, Y YA ESTÁ, ¿QUÉ TE IBA A DECIR?, Y ESO, Y TAL, ¡QUÉ VA!, DESDE LUEGO, PUES ESO, HASTA LUEGO, ¿QUÉ HAY?. ¿QUÉ TAL?. ¿QUÉ PASA?, ¿CÓMO TE VA LA VIDA?, ¿QUÉ TE CUENTAS?, ¡QUÉ BONITO!, ¿QUÉ ME DICES?, ¡QUÉ FUERTE!, ¡QUÉ PASADA!, ¡QUÉ PUTADA!, ¡HIJA MÍA!, PUES VAYA, ¿QUÉ QUIERES QUE TE DIGA?

Para ello, hemos llevado a cabo dos tipos de análisis:

1. Un análisis de autoría a través de la prueba chí cuadrado, con la que hemos comparado textos aparentemente dubitados con textos indubitados y con un conjunto de textos generales para comprobar si el programa es eficaz para nuestro corpus.

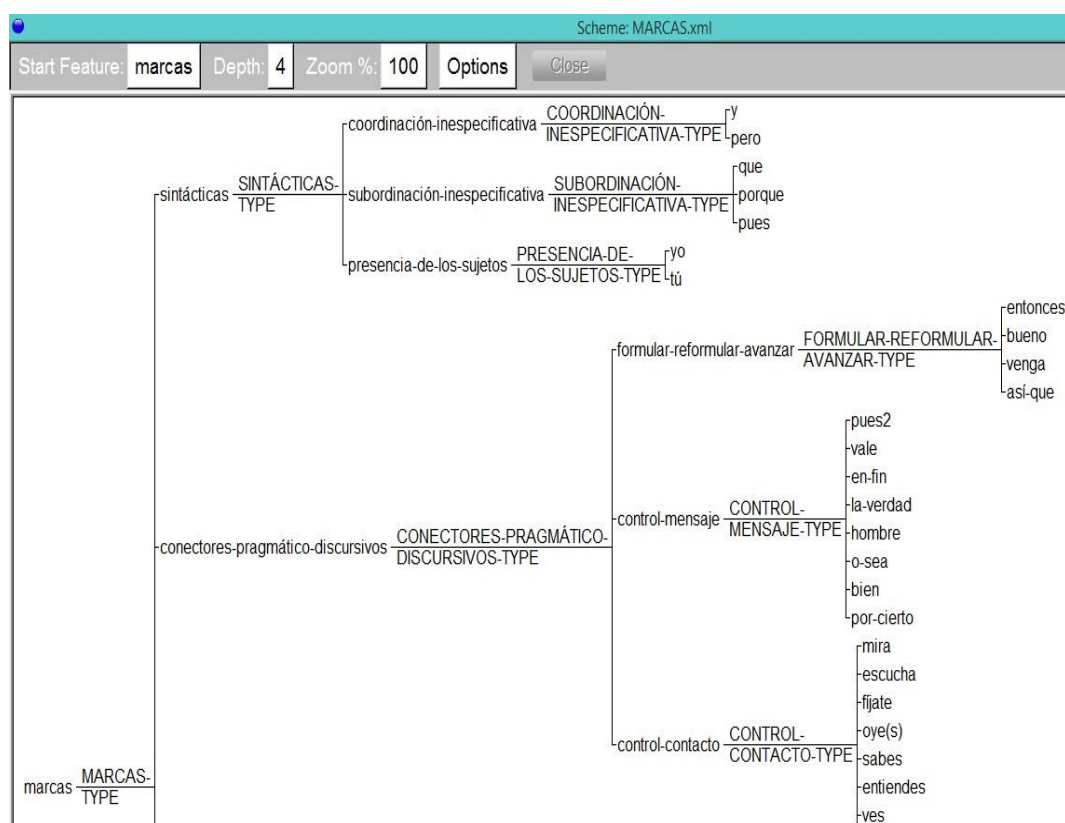
2. A partir de la frecuencia de uso de las marcas seleccionadas, estableceremos perfiles de uso con respecto a las características de los autores seleccionados.

6.2.1. Análisis Chí cuadrado:

Primeramente, realizamos un esquema en forma de árbol en el programa UAM Corpus Tool para poder anotarlo posteriormente y obtener así los resultados.

Aquí adjuntamos una pequeña imagen del esquema en forma de árbol que hemos configurado en el programa para poder anotar posteriormente nuestras marcas:

Esquema marcas registro coloquial en UAM Corpus Tool



Los datos que se extraen como producto de la aplicación de Chi Cuadrado permiten establecer conclusiones preliminares con respecto a la autoría de los textos analizados. A continuación, desarrollaremos el análisis pormenorizado de cada uno de ellos.

Tabla 1: Comparación de las diferencias estadísticas del Autor 1 con la herramienta Chi Cuadrado:

Dubitado 1 – Indubitado 1	Dubitado 1 – General	Indubitado 1 – General
Diferencias encontradas: 2	Diferencias encontradas: 8	Diferencias encontradas: 5
Expresiones subjetividad: 6.417 +++		Expresiones subjetividad: 10.640 +++
Expresiones atenuadoras: 6.848 +++	Expresiones atenuadoras: 14.144 +++	

	Sintácticas: 10.207 +++	
	Y: 6.039 +++	
	Signos exclamación: 11.722 +++	
	Signos puntuación: 6.645 +++	
	Emoticonos: 18.302 +++	Emoticonos: 17.772 +++
	Creo: 6.976 +++	
	Me parece: 16.614 +++	
		Fórmulas rutinarias: 6.371 +++
		En fin: 17.775 +++
		Bien: 7.112 +++

La comparación del texto dubitado con el indubitado arroja solamente dos diferencias y ambas de carácter general (no se asocian con una marca concreta sino con grupos de marcas). Esto nos permite afirmar que ambos textos (dubitado e indubitado) pertenecen o bien al mismo autor o bien a autores de rasgos bastante similares en el registro y entorno comunicativos de los que hablamos.

Cuando enfrentamos el texto dubitado con el general encontramos que solamente aparecen ocho diferencias destacables. Sin embargo, dicho número debe entenderse de manera relativa en la medida en que el programa incluye tanto datos de grupos de marcas (caso de “Expresiones atenuadoras” o “Sintácticas”) como aquellos que se relacionan con marcas concretas (“Y”, “Creo”, “Me parece”). Dado que algunas de las marcas concretas pertenecen a los grupos de marcas que también se incluyen en el análisis, los datos deben considerarse, como decimos, de manera relativa. Esto se traduce en que, probablemente, las diferencias entre el texto dubitado y el general sean menores. Lo mismo sucede al enfrentar el texto indubitado con el grupo general. Si comparamos las diferencias que el programa identifica en este caso con las halladas en la comparación entre el dubitado y el general, puede parecer que el indubitado manifiesta menos diferencias. Teniendo en cuenta nuestro comentario sobre la mezcla que el programa lleva a cabo entre marcas y grupos de marcas, podemos afirmar que las distinciones generales dubitado-general e indubitado-general son bastante similares en cuanto al número.

Tabla 2: Comparación de las diferencias estadísticas del Autor 2 con la herramienta Chí Cuadrado:

Dubitado 2 – Indubitado 2	Dubitado 2 - General	Indubitado 2 - General
Diferencias encontradas: 3	Diferencias encontradas: 7	Diferencias encontradas: 5
Fórmulas rutinarias: 10.121 +++	Fórmulas rutinarias: 15.271 +++	
La verdad: 5.653 +++		La verdad: 57.142 +++
Y eso: 5.713 +++	Y eso: 16.736 +++	
	Expresiones subjetividad: 15.098 +++	Expresiones subjetividad: 30.153 +++
	Venga: 11.959 +++	
	Oye(s): 21.267 +++	
	Signos puntuación: 7.772 +++	Signos puntuación: 9.175 +++
	¡Qué bonito!: 73.992 +++	
		Signos exclamación: 5.882 +++
		Emoticonos: 13.279 +++

La comparación del texto dubitado con el indubitado arroja solamente tres diferencias, una de carácter general (“Fórmulas rutinarias”) y otras dos relacionadas con marcas específicas (“La verdad”, “Y eso”). Esto nos permite afirmar que, de nuevo, ambos textos (dubitado e indubitado) pertenecen o bien al mismo autor o bien a autores de rasgos bastante similares en el registro y entorno comunicativos de los que hablamos.

La comparación entre el texto dubitado y el grupo general pone de manifiesto únicamente siete diferencias encontradas. Como en el caso del texto dubitado 1, algunas pertenecen al ámbito de los grupos de marcas (“Fórmulas rutinarias”, “Expresiones de subjetividad”, “Signos de puntuación”) mientras que otras son marcas concretas (“Venga”, “Oye”, “Y eso”). De ahí que podamos volver a trabajar con la hipótesis de que las diferencias entre indubitado

y grupo general sean menores de las que el programa indica. Al contrastar el texto indubitado con el grupo general en este segundo caso el número de diferencias resulta, como ya ocurría en el caso 1, es bastante reducido y similar al que arrojan los datos de la comparación dubitado 2 - general. La coincidencia relativa entre el número de diferencias de la comparación dubitado2-general e indubitado2-general permiten trabajar con la hipótesis de que ambos textos han sido producidos por el mismo autor o, al menos, por autores de rasgos lingüísticos muy similares. Esta idea toma mayor consistencia cuando, en el análisis interno de las diferencias expuestas en la tabla correspondiente, comprobamos cómo no solamente se trata de un número similar de diferencias sino incluso de diferencias específicas que se repiten en ambos textos con respecto al general (“Expresiones de subjetividad”, “Signos de puntuación”). Como dato curioso en este segundo grupo de análisis debemos destacar que, mientras que en el análisis de dubitado1-indubitado1 frente al grupo el empleo de emoticonos destacaba como marca distintiva, en este caso no ocurre (se registra la marca en el texto dubitado 2 mientras que en el indubitado 1 no aparece).

Tabla 3: Comparación de las diferencias estadísticas del Autor 3 con la herramienta Chí Cuadrado:

Dubitado 3 – Indubitado 3	Dubitado 3 - General	Indubitado 3 - General
Diferencias encontradas: 2	Diferencias encontradas: 7	Diferencias encontradas: 10
Expresiones atenuadoras: 6.449 +++	Expresiones atenuadoras: 5.476 +++	
Vale: 7.915 +++	Signos puntuación: 13.060 +++	Signos puntuación: 20.450 +++
	Coordinación inespecificativa: 6.584 +++	Coordinación inespecificativa: 7.813 +++
	Subordinación inespecificativa: 38.844 +++	Subordinación inespecificativa: 61.964 +++

	Que: 33.859 +++	Que: 54.154 +++
	Pienso: 5.753 +++	
	Emoticonos: 36.505 +++	Emoticonos: 78.463 +++
		Porque: 7.230 +++
		Tú: 11.890 +++
		Control mensaje: 10.681 +++
		Y nada: 25.842 +++
		Signos exclamación: 7.372 +++

La comparación del texto dubitado con el indubitado arroja solamente dos diferencias, una de carácter general (“Expresiones atenuadoras”) y otra relacionada con marcas específicas (“Vale”). En este caso la idea de que ambos textos (dubitado e indubitado) pertenecen o bien al mismo autor o bien a autores de rasgos bastante similares en el registro y entorno comunicativos de los que hablamos aparece reforzada por un rasgo ya comentado en los análisis de los grupos de textos anteriores: la diferencia “Expresiones atenuadoras” puede perfectamente contener los casos de “Vale” que el programa singulariza como marca específica destacada, con lo que las diferencias reales entre las marcas distintivas de ambos textos (dubitado e indubitado) son realmente menores.

De nuevo en el ámbito de la distinción dubitado-general e indubitado-general los resultados apuntan en la misma dirección que los de grupos anteriores. Así, las diferencias que el programa detecta entre el texto dubitado 3 y el grupo general son siete, mientras que las halladas entre el texto indubitado y dicho grupo general son diez. Si bien el número de diferencias vuelve a ser bastante similar, volvemos a advertir de que probablemente sea posible reducirlo. Diferencias como “Subordinación inespecificativa” contienen a otras como “Que”, estando ambas singularizadas y contabilizadas de manera independiente en el total (caso de la comparación dubitado 3 - general). Exactamente el mismo caso se da en el par indubitado 3 - general y con las mismas marcas. Con respecto al empleo de los emoticonos, en ambos pares de comparación (dubitado 3 - general e indubitado 3 - general) la marca aparece como distintiva frente al grupo. Todos los datos analizados permiten trabajar con la hipótesis de que, o bien el texto dubitado 3 y el indubitado 3 pertenecen al mismo emisor, o

bien han sido articulados por emisores de un perfil lingüístico bastante similar con respecto a las marcas seleccionadas para nuestro análisis.

Tabla 4: Comparación de las diferencias estadísticas del Autor 4 con la herramienta Chí Cuadrado:

Dubitado 4 – Indubitado 4	Dubitado 4 - General	Indubitado 4 - General
Diferencias encontradas: 5	Diferencias encontradas: 10	Diferencias encontradas: 13
Conectores pragmáticos: 12.604 +++		Conectores pragmáticos: 10.113 +++
Fórmulas rutinarias: 8.027 +++		Fórmulas rutinarias: 6.181 +++
Coordinación inespecificativa: 19.000 +++	Coordinación inespecificativa: 11.004 +++	Coordinación inespecificativa: 12.875 +++
	Subordinación inespecificativa: 21.463 +++	Subordinación inespecificativa: 32.143 +++
Control mensaje: 11.389 +++		Control mensaje: 11.513 +++
Bien: 7.041 +++		
	Expresiones subjetividad: 25.235 +++	Expresiones subjetividad: 43.205 +++
	Que: 20.637 +++	Que: 45.128 +++
	Control contacto: 7.754 +++	
	Hombre: 13.733 +++	
	O sea: 48.639 +++	O sea: 84.984 +++
	Sabes: 14.955 +++	

	Signos puntuación: 15.069 +++	Signos puntuación: 38.697 +++
	Emoticonos: 25.860 +++	Emoticonos: 34.526 +++
		Y: 5.945 +++
		Vale: 10.058 +++
		Signos exclamación: 7.224 +++

Cuando comparamos el texto dubitado 4 con el indubitado 4 el número de diferencias que el programa refleja entre ambos llama la atención con respecto al análisis de resultados en los grupos anteriores, siendo cinco las diferencias registradas. Sin embargo, lo que en un principio puede tomarse como un rasgo negativo en lo que a la atribución de autoría respecta, no lo es si consideramos las diferencias concretas de que se trata: cuatro de las cinco señaladas se corresponden con grandes grupos de marcas (“Conectores pragmáticos”, “Fórmulas rutinarias”, “Coordinación inespecificativa”, “Control del mensaje”). esto implica que las diferencias entre ambos textos pueden traducirse en la ausencia o presencia de un conector o de una fórmula concretos, pero nunca de forma tan reiterada que permita evidenciarlo como rasgo distintivo propio. De ahí la relatividad con la que debemos tomar la identificación de estas diferencias. Podemos, como en casos anteriores, afirmar el texto dubitado 4 frente al indubitado 4 no manifiesta diferencias significativas en lo que respecta a las marcas analizadas.

En la segunda parte de nuestro análisis (comparación entre dubitado 4 - grupo general e indubitado 4 - grupo general) volvemos a encontrar evidencias ya registradas. Así, el número de diferencias entre dubitado 4 - indubitado 4 no resulta significativo (en la comparación entre dubitado 4 - grupo general el programa registra diez, mientras que en el análisis de los datos del grupo indubitado 4 - grupo general el cómputo de diferencias es de trece). En suma, no se identifica un número de diferencias que permita distanciar entre sí la autoría del texto dubitado 4 frente al indubitado 4. Por otra parte, y de nuevo reproduciendo análisis de los rasgos identificados en grupos anteriores, parece posible reducir el número de diferencias relativas en cada uno de los grupos (dubitado 4 - grupo general, indubitado 4 - grupo general) en la medida en que un buen número de ellas aluden a grupos de marcas en las que se incluyen las marcas específicas singularizadas como distintivas frente al grupo general en cada caso. Se trata de diferencias como “Coordinación inespecificativa”, “Subordinación inespecificativa”, “Expresiones de subjetividad”, “Control del contacto” (comparación texto

dubitado 4 - grupo general) o, en el caso de la comparación entre el texto indubitado 4 y el general, diferencias como “Conectores pragmáticos”, “Fórmulas rutinarias”, “Coordinación inespecificativa”, “Subordinación inespecificativa”, “Control del mensaje” y “Expresiones subjetividad”. Con respecto al empleo de emoticonos como marca distintiva, vuelve a ser diferencia común en la comparativa de ambos textos (dubitado e indubitado) con el grupo general. En suma, la comparación de las diferencias relativas de cada texto (dubitado e indubitado) con el grupo general permiten acercar, de nuevo, la autoría de ambos textos.

Tabla 5: Comparación de las diferencias estadísticas del Autor 5 con la herramienta Chí Cuadrado:

Dubitado 5 – Indubitado 5	Dubitado 5 - General	Indubitado 5 - General
Diferencias encontradas: 1	Diferencias encontradas: 3	Diferencias encontradas: 9
	Expresiones subjetividad: 16.798 +++	Expresiones subjetividad: 27.883 +++
Entonces: 10.246 +++	Entonces: 11.494 +++	
	Emoticonos: 10.014 +++	Emoticonos: 36.839 +++
		Conectores pragmáticos: 15.079 +++
		Que: 6.053 +++
		Porque: 10.449 +++
		Control mensaje: 15.964 +++
		Vale: 7.508 +++
		Entiendes: 11.446 +++
		Signos exclamación: 12.655 +++

Si en grupos de textos anteriormente analizados la proximidad en la autoría resultaba fácil de identificar, en la comparación del texto dubitado 5 con el indubitado 5 alcanzamos el nivel de

mayor coincidencia: el programa solamente registra una diferencia significativa y, además, bastante más específica que algunas de las anteriores. Se trata de la diferencia marcada por la fórmula rutinaria de carácter conclusivo “Entonces”. Como decimos, los niveles de coincidencias entre los textos analizados resultan elevados.

La comparación entre los grupos texto dubitado 5 - grupo general por una parte y texto indubitado 5 - grupo general por otra parece distanciar, frente a los datos de la comparativa entre ambos textos, sus posibilidades de autoría común (en el grupo texto dubitado 5 - grupo general se registran tres diferencias, mientras que en el grupo texto indubitado 5 - grupo general suben hasta un total de nueve). Como en ocasiones anteriores, lo que puede parecer un dato negativo con respecto a la atribución de autoría no lo es tanto si tenemos en cuenta que cinco de las nueve diferencias identificadas por el programa en la comparación texto indubitado 5 - grupo general se refieren a grupos de marcas y no a rasgos concretos (“Expresiones subjetividad”, “Emoticonos”, “Conectores pragmáticos”, “Control del mensaje”, “Signos de exclamación”). En definitiva, podemos acercar la autoría de los textos dubitado 5 e indubitado 5 como hipótesis de trabajo pendiente de un análisis de mayor calado.

La lectura de los datos que el programa ofrece para la atribución de autoría de los textos seleccionados nos permite, además del dictamen individual esbozado en líneas anteriores, concluir que:

- 1) En todos los casos la autoría del texto dubitado aparece como bastante cercana a los utilizados como textos indubitados. Si tenemos en cuenta que, como señalábamos en la introducción a la parte práctica de nuestro trabajo, los textos que hemos señalado como dubitados pertenecen a los mismos autores que los textos indubitados de cada grupo, los resultados nos permiten atestiguar el funcionamiento eficaz del programa con respecto al análisis de marcas propuesto.
- 2) La diferencia existente entre el volumen de texto de los textos que hemos singularizado frente al del grupo general autoriza a tomar los datos obtenidos y las conclusiones esbozadas como absolutamente preliminares, pendientes de refrendar mediante el cotejo con texto de mayores dimensiones.

6.2.2. Tabla de error:

En la siguiente tabla se comparan los textos dubitados con los indubitados para observar su comportamiento. Los datos señalados en verde nos indican que, efectivamente los textos

producidos por el mismo autor poseen menos diferencias que si son escritos por autores diferentes. Es por ello por lo que las diferencias entre los textos del mismo individuo son menores que las diferencias que existen entre textos de diferentes personas. Los datos señalados en rojo muestran los casos en los que el programa da error en nuestra investigación. En nuestro caso, hemos observado que al comparar el dubitado 5 con el indubitado 1 solo se ha encontrado 1 diferencia, cuando se supone que debe haber más de las que se dan entre el dubitado 5 y el indubitado 5. Lo mismo ocurre al comparar el dubitado 5 con el indubitado 2. Una posible razón por la que habrá ocurrido esto es porque no tenemos suficiente texto del autor 5 para poder realizar este estudio. Sin embargo, podemos concluir que nuestro porcentaje de acierto es del 92%.

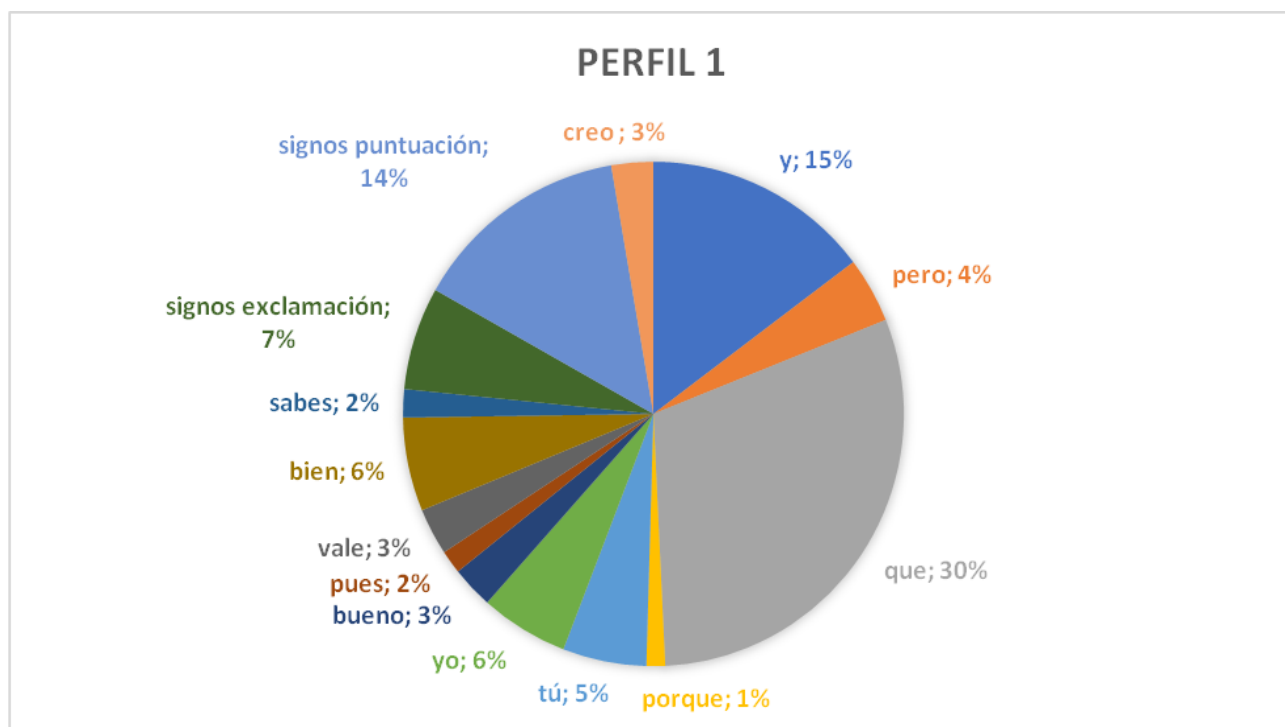
	Indubitado 1	Indubitado 2	Indubitado 3	Indubitado 4	Indubitado 5
Dubitado 1	2	7	11	13	9
Dubitado 2	10	3	15	15	13
Dubitado 3	8	6	2	7	8
Dubitado 4	11	11	10	5	15
Dubitado 5	1	2	10	10	1

6.2.3. Análisis perfiles lingüísticos:

En el apartado anterior de nuestro estudio hemos analizado la manera en que los textos seleccionados como dubitados podían analizarse en términos de atribución de autoría con respecto a otros de grupos y autores prefijados. En adelante, trazaremos un perfil lingüístico de cada uno de los autores analizados basándonos en el empleo de las marcas que hemos tomado como base. Dado que el análisis de autoría nos llevó a la conclusión preliminar de que en todos los casos el texto dubitado podía corresponder al autor del texto con el que se comparaba, podemos utilizar el empleo de marcas en ambos tipos de texto (dubitado e indubitado) para trazar cada perfil. A continuación, presentaremos dichos perfiles donde se puede observar el total de marcas dividido por tipos y el porcentaje de uso de cada una de ellas.

Perfil 1 - Características: Hombre, 58 años, estudios universitarios.

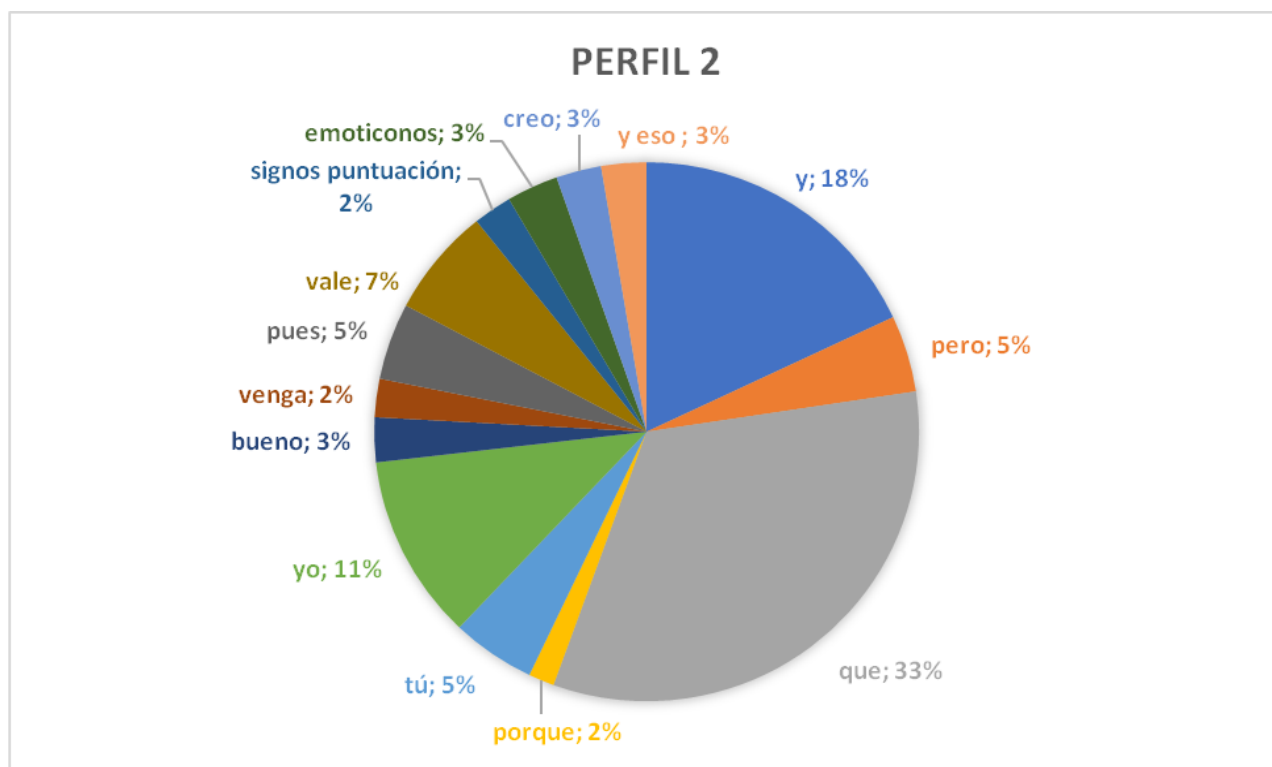
Gráfico 1 – Perfil lingüístico 1:



Como podemos observar en el gráfico, las marcas que el perfil 1 emplea con mayor frecuencia en sus textos son: el nexos subordinante inespecificativo “que” con un 30%, el nexos coordinante inespecificativo “y” con un 15% y los signos de puntuación con un 14%.

Perfil 2 - Características: Hombre, 23 años, estudios universitarios.

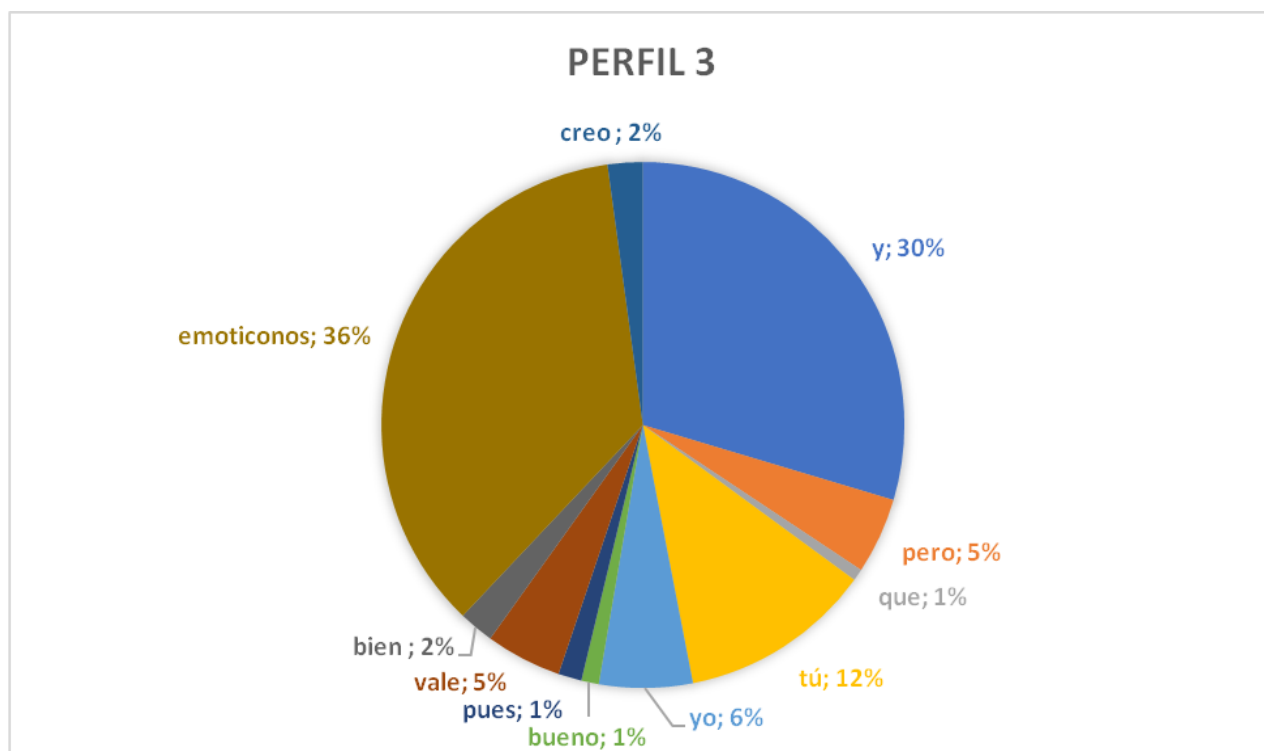
Gráfico 2 – Perfil lingüístico 2:



El perfil 2 utiliza las siguientes marcas en sus textos con mayor frecuencia: el nexos subordinante inespecificativo “que” con un 33%, el nexos coordinante inespecificativo “y” con un 18% y el sujeto de la enunciación “yo” con un 11%.

Perfil 3 - Características. Mujer, 46 años, estudios universitarios.

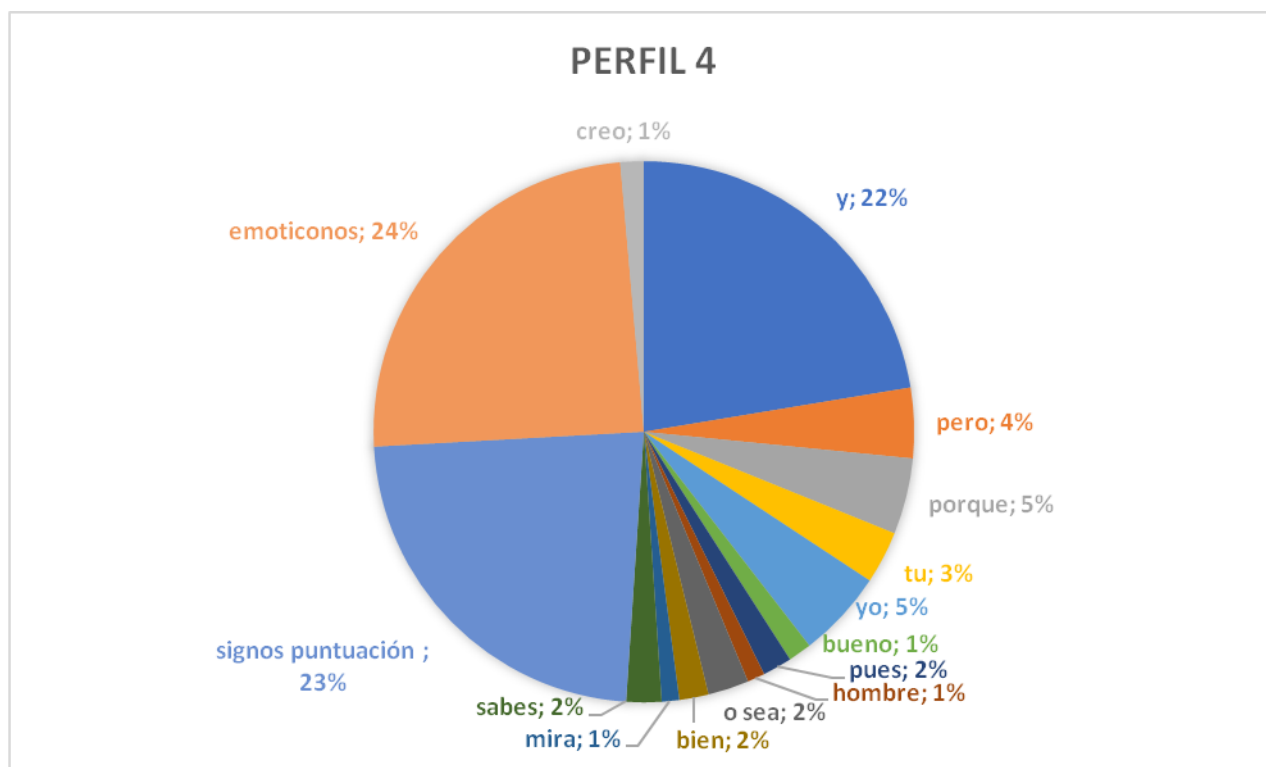
Gráfico 3 – Perfil lingüístico 3:



El perfil 3 usa las mayoritariamente estas marcas en sus textos: emoticonos, nexos coordinantes inespecíficos “y” y el sujeto de la enunciación “tú” con un 36%, 30% y 12% respectivamente.

Perfil 4 - Características: Mujer, 34 años, estudios universitarios.

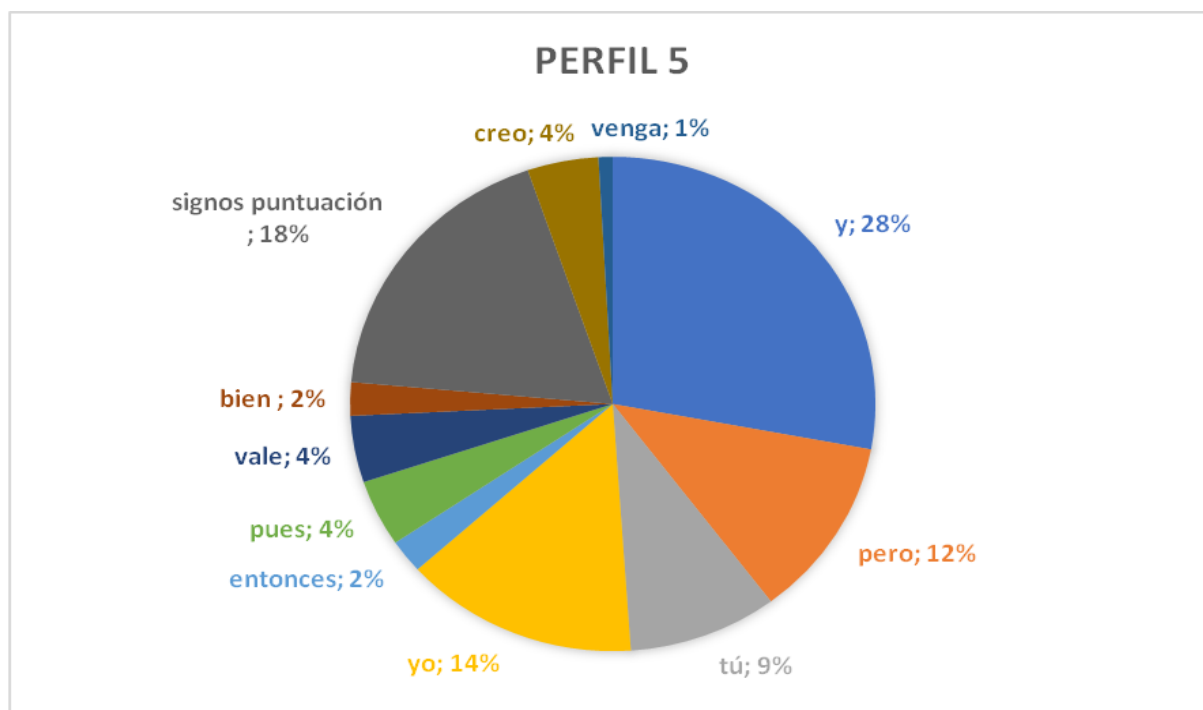
Gráfico 4 – Perfil lingüístico 4:



Este perfil, como vemos en el gráfico, tiende a utilizar en mayor medida en sus textos las marcas: emoticonos, con un 24%, signos de puntuación, con un 23% y el nexos coordinante inespecificativo “y” con un 22%.

Perfil 5 - Características: Hombre, 22 años, estudios universitarios.

Gráfico 5 – Perfil lingüístico 5:



Este perfil emplea con mayor frecuencia en sus textos las siguientes marcas lingüísticas: el nexos coordinante inespecificativo “y” con un 28%, los signos de puntuación con un 18% y el sujeto de la enunciación “yo” con un 14%.

Los rasgos lingüísticos compartidos por los cinco hablantes son: los nexos coordinantes inespecificativos “y” y “pero”, el nexos subordinante inespecificativo “pues”, el conector pragmático “bueno” y la expresión atenuadora “creo”, Es por ello por lo que no los incluiremos en nuestras tablas comparativas ya que, al estar presentes en todos los perfiles, parecen ser rasgos insuficientes para discriminar distintos perfiles de hablantes. Aunque los sujetos de la enunciación “yo” y “tú” también aparecen en todos los perfiles, consideramos que es pertinente incluirlos en las tablas comparativas porque sus porcentajes de uso son diferentes en hombres y mujeres. Los hombres los suelen emplear en torno a un 10,3% (yo) y 6,3% (tú) de media y sin embargo, las mujeres usan los sujetos de la enunciación yo y tú un 5,5% y un 7,5% respectivamente.

Las edades de los hablantes analizados se distribuyen en **edad madura** (58, 46), **jóvenes** (23, 22) y finalmente **adultos** (34). Además, de los cinco autores analizados, tres son hombres y dos son mujeres.

Tabla 1 - Rasgos compartidos por los hombres:

58 años (Perfil 1):	23 años (Perfil 2):	22 años (Perfil 5):
Que 30 %	Que 33%	Que 0%
Signos puntuación 14 %	Signos puntuación 2%	Signos puntuación 18%
Yo 6%	Yo 11%	Yo 14%
Tú 5%	Tú 5%	Tú 9%
Vale 3%	Vale 7%	Vale 4%
Venga 0%	Venga 2%	Venga 1%

Como podemos comprobar en la tabla, estos son los rasgos que podrían utilizarse en la configuración los perfiles de los hombres que hemos analizado en el estudio. Es destacable la marca “que” pues los hombres (salvo el último) la utilizan con mucha frecuencia en sus textos, en concreto un 30% y un 33%. Esta marca, por otro lado, no es usada ninguna vez por las mujeres, por lo que podría entenderse como un indicador de sexo masculino. La marca “venga” solo es utilizada por los hombres jóvenes mientras que el hablante mayor y las mujeres no la usan ninguna vez en sus textos producidos en wasap. El uso de los sujetos de la enunciación “yo” y “tú” tiende a ser mayor en los mensajes producidos por los hombres que en los de las mujeres, por lo que también sería una marca a tener en cuenta a la hora de la identificación de hablantes. El conector pragmático discursivo “vale” aparece en los mensajes de los tres hombres mientras que solo una de las mujeres la usa también por lo que se podría pensar que es una marca más frecuente en hombres que en mujeres. Finalmente, la marca “signos de puntuación” aparece más frecuentemente en hombres que en las mujeres (solo el perfil 4 los usa) por lo que podría llegar a ser un buen rasgo identificador del sexo masculino.

Tabla 2 - Rasgos compartidos por las mujeres:

46 años (Perfil 3):	34 años (Perfil 4):
Emoticonos: 36%	Emoticonos: 24%
Yo 6%	Yo 5%
Tú 12%	Tú 3%
Bien 2%	Bien 2%

Esta tabla nos muestra los rasgos que ambas mujeres poseen a la hora de producir mensajes en la aplicación wasap. El dato más interesante que hemos encontrado es que las dos emplean emoticonos en sus textos a diferencia de los hombres, ya que ninguno (salvo el perfil número 2 que sí los emplea, pero tan solo un 3%) los utiliza a la hora de hablar por wasap. Los demás rasgos mostrados en la tabla se distribuyen de manera similar en los textos de los dos perfiles por lo que podrían ser indicadores del sexo femenino. Otro dato curioso es que las mujeres no emplean la marca “que” en sus textos mientras que los hombres sí lo hacen. Además, las mujeres seleccionadas para el estudio sí han empleado el conector pragmático discursivo “bien” mientras que los hombres (salvo el perfil 1) no lo han hecho.

Ahora compararemos las marcas por rangos de edad. Como solo hemos escogido a un hablante del grupo que hemos denominado “adultos” para nuestra investigación, este perfil no será incluido en las siguientes tablas.

Tabla 3 - Rasgos compartidos por los hablantes de **edad madura**:

Hombre 58 años (Perfil 1):	Mujer 46 años (Perfil 3):
Yo 6%	Yo 6%
Tú 5%	Tú 12%
Vale 3%	Vale 5%
Que 30%	Que 1%

Como podemos observar en esta tabla, estos dos hablantes de edades próximas emplean la marca “yo” y “vale” con una frecuencia similar en sus mensajes. Sin embargo, difieren en el uso del “que” y del “tú”.

Tabla 4 - Rasgos compartidos por los hablantes jóvenes:

23 años (Perfil 2):	22 años (Perfil 5):
Que 33%	Que 0%
Signos puntuación 2%	Signos puntuación 18%
Yo 11%	Yo 14%
Tú 5%	Tú 9%
Vale 7%	Vale 4%
Venga 2%	Venga 1%

Un rasgo destacable que puede ser indicador de la diferencia de edad entre el hombre de edad madura y entre los otros dos más jóvenes es la marca “yo”. El hablante de 58 la emplea un 6% mientras que los hablantes jóvenes la usan en sus textos un 11% y un 14%. Lo contrario ocurre con la marca “que”. El hombre de edad madura la usa un 30% mientras que el joven de 22 años no la usa (exceptuando al de 23 años que sí la utiliza). Ambos hablantes jóvenes emplean “vale” y venga” con una frecuencia similar.

7. Conclusión:

- 1) Tanto en el apartado dedicado al análisis chí cuadrado de los datos obtenidos con el programa UAM Corpus Tool como en el dedicado al esbozo de los perfiles de hablantes, las conclusiones alcanzadas deben entenderse como absolutamente preliminares por varias razones:
 - a. El volumen de la muestra utilizada no permite entender los datos de manera absoluta.
 - b. El número de informantes seleccionados para el trabajo debe ser ostensiblemente ampliado de cara al afianzamiento de las conclusiones alcanzadas
 - c. El corpus utilizado carece de especificaciones de uso que, desde una perspectiva pragmática, contribuirían a clarificar el empleo de algunas de las marcas.

- 2) Si bien las objeciones detalladas en el primer apartado de nuestras conclusiones limitan el alcance final de nuestro trabajo, sí es posible establecer, como decimos, algunas conclusiones preliminares:
- a. El entorno comunicativo de wasap mantiene el uso de muchas de las marcas consideradas como pertenecientes al registro coloquial en español, al mismo tiempo que incluye otras nuevas (ej.: emoticonos, signos de puntuación).
 - b. El análisis de los datos obtenidos permite afirmar que el programa UAM Corpus Tool mantiene un funcionamiento exitoso en estudios de este tipo (entorno wasap, registro coloquial).
 - c. Las semejanzas y diferencias manifestadas entre los informantes permiten trazar perfiles de uso con respecto a las marcas elegidas para nuestra investigación.

Todas estas conclusiones deberán ser ratificadas en una futura ampliación de nuestra investigación que permita entender como estables o, en su caso, modificar como sea oportuno los perfiles lingüísticos trazados. Consideramos que estudios de este tipo ayudan a la consolidación de disciplinas como la lingüística forense como instrumentos eficaces en la investigación criminal.

8. Bibliografía:

- Alvarado Ortega, M.B. (2010). *Las fórmulas rutinarias del español: teoría y aplicaciones*. Frankfurt am Main: Peter Lang.
- Beinhauer, W. (1978). *El español coloquial*. (1964). Madrid: Editorial Gredos.
- Briz Gómez, A. (1998). *El español coloquial: Situación y uso*. (1996). Madrid: Arco Libros.
- Briz Gómez, A. (2001). *El español coloquial en la conversación. Esbozo de pragmatogramática*. (1998). Barcelona: Ariel Lingüística.
- Briz Gómez, A. (2010). *El español coloquial: Situación y uso*. (1996). Madrid: Arco Libros.
- Campos, Carrasco, N. (2018). Cortesía y grado de fuerza en la realización de los “actos amenazadores de la imagen” en español. En XXVI (Ejemplar dedicado a: Pragmática intercultural, social y cognitiva), págs. 57-78.

- Cicres, J., Turell, M. T. (2014). Investigación, docencia y práctica profesional en Lingüística Forense. En E. Garayzábal. M. Jiménez., M. Reigosa. *Lingüística forense: la lingüística en el ámbito Legal y Policial*. Madrid: Euphonia Ediciones.
- Crespo Miguel, M. (2018). “Retos de la estilística forense en el ámbito del discurso electrónico delictivo”. Cádiz: Universidad de Cádiz.
- Criminalística*. Guardia Civil (s.f.). Recuperado 15 mayo, 2019, de <http://www.guardiacivil.es/es/institucional/Conocenos/especialidades/InvestigacionCientifica/index.html>
- Fitzgerald, J. R. (2014). Atribución de autoría y supuestas notas de suicidio: Análisis lingüístico forense y su papel en los tribunales penales estadounidenses en dos crímenes violentos ocurridos en 2007. En E. Garayzábal. M. Jiménez., M. Reigosa. *Lingüística forense: la lingüística en el ámbito Legal y Policial*. Madrid: Euphonia Ediciones.
- ForensicLab - UVAL (UPF). (s.f.). Recuperado 15 mayo, 2019, de <https://www.upf.edu/es/web/uval/forensiclab>
- García Picornell, I. (2014). La aplicación de la atribución de autoría en la investigación e inteligencia: La aplicación práctica (y su problemática). En E. Garayzábal. M. Jiménez., M. Reigosa. *Lingüística forense: la lingüística en el ámbito Legal y Policial*. Madrid: Euphonia Ediciones.
- Grupo Académico ILFE*. (s.f.). Recuperado 15 mayo, 2019, de <https://grupoacademicoilfe.weebly.com/>
- Instituto de Lingüística Aplicada – ILA*. (s.f.). Recuperado 15 mayo, 2019, de <https://ila.uca.es/>
- Jiménez Bernal, M., Reigosa Riveiros, M., Garayzábal Heinze, E. (2014). La lingüística forense: Licencia para investigar la lengua. En E. Garayzábal. M. Jiménez., M. Reigosa. *Lingüística forense: la lingüística en el ámbito Legal y Policial*. Madrid: Euphonia Ediciones.
- López Serena, A. (2007). El concepto de ‘español coloquial’: vacilación terminológica e indefinición del objeto de estudio, *Oralia. Volumen*. (10). Universidad de Extremadura / Universidad de Sevilla: Arco Libros.
- McMenamim, G. R. (2002). *Forensic Linguistics, Advances in Forensic Stylistics*. United States of America: CRC Press.
- Narbona Jiménez, A. (2015). *Sintaxis del español coloquial*. Sevilla: Editorial Universidad de Sevilla.

- Olsson, J. (2008). *Forensic Linguistics: Second edition*. Great Britain: Continuum International Publishing Group.
- Página oficial de la DGP-Comisaría General de Policía Científica. (s.f.). Recuperado 15 mayo, 2019 de https://www.policia.es/org_central/cientifica/servicios/tp_acustic_foren.html
- Pascual, F. J. B., Urbón, C. R. (2009). Valuación y cuantificación de algunas técnicas de Atribución de autoría en textos españoles. *Castilla: Estudios de Literatura*, 27-47.
- Qué es una prueba de hipótesis* - Minitab. (s.f.). Recuperado 1 mayo, 2019, de <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/basics/what-is-a-hypothesis-test/>
- Qué es una prueba de hipótesis* (2019). Recuperado 1 mayo, 2019, de <https://www.addlink.es/noticias/minitab/2852-que-es-una-prueba-de-hipotesis>
- Queralt, S. (2014). Acerca de la prueba lingüística en atribución de autoría hoy, *Revista de Llengua i Dret. Volum* (62).
- Queralt, S. (s.f.). *Sheila Queralt Home*. Recuperado 15 mayo, 2019, de <https://www.sq-linguistasforenses.com/sheilaqueralt.html>
- Ramírez Salado, M. (2017). Antecedentes de la lingüística forense: ¿desde cuándo se estudia el lenguaje como evidencia?, *Pragmalingüística. Volumen* (25), 525-539.
- Sampietro, A. (2016). *Emoticonos y emojis. Análisis de su historia, difusión y uso en la comunicación digital actual*. Valencia: Facultat de Filologia, Traducció i Comunicació, Universitat de València. (<http://infoling.org/repository/PhDdiss-Infoling-83-5-2016.pdf>)
- Sánchez Esteban, M. (2016). *Lingüística Forense Básica, Metodología para la atribución de autoría de un texto*. Valencia: Tirant lo Blanch.
- Stefanova Spassova, M. (2009). *El potencial discriminatorio de las secuencias de categorías gramaticales en la atribución forense de autoría de textos en español* (tesis doctoral), Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Svartvik, J. (1968). *The Evans Statements, a case for Forensic Linguistics*. Gotemburgo: Gothenburg studies in English, University of Gothenburg.
- Triola, M. F. (2009). *Estadística*. México: Pearson Educación.

Turell, M. T. (2005). *Lingüística Forense, Lengua y Derecho: conceptos, métodos y aplicaciones*. Barcelona: Institut Universitari de Lingüística Aplicada Universitat Pompeu Fabra.

Vigara Tauste, A. M. (2005). *Morfosintaxis del español coloquial. Esbozo estilístico*. (1992), Madrid: Gredos.